Title

Brain aging patterns in a large and diverse cohort of 49,482 individuals.

Permalink

https://escholarship.org/uc/item/3j47g1md

Journal

Nature Medicine, 30(10)

Authors

Yang, Zhijian
Wen, Junhao
Erus, Guray
et al.

Publication Date

2024-10-01

DOI

10.1038/s41591-024-03144-x

Published in final edited form as:
Nat Med. 2024 October ; 30(10): 3015–3026. doi:10.1038/s41591-024-03144-x.

Brain aging patterns in a large and diverse cohort of 49,482 individuals

A full list of authors and affiliations appears at the end of the article.

Abstract

Brain aging process is influenced by various lifestyle, environmental and genetic factors, as well as by age-related and often coexisting pathologies. Magnetic resonance imaging and artificial intelligence methods have been instrumental in understanding neuroanatomical changes that occur during aging. Large, diverse population studies enable identifying comprehensive and representative brain change patterns resulting from distinct but overlapping pathological and biological factors, revealing intersections and heterogeneity in affected brain regions and clinical phenotypes. Herein, we leverage a state-of-the-art deep-representation learning method, Surreal-GAN, and present methodological advances and extensive experimental results elucidating brain aging heterogeneity in a cohort of 49,482 individuals from 11 studies. Five dominant patterns of brain atrophy were identified and quantified for each individual by respective measures, R-indices. Their associations with biomedical, lifestyle and genetic factors provide insights into the etiology of observed variances, suggesting their potential as brain endophenotypes for genetic and lifestyle risks. Furthermore, baseline R-indices predict disease progression and mortality, capturing early changes as supplementary prognostic markers. These R-indices establish a dimensional approach to measuring aging trajectories and related brain changes. They hold promise for precise diagnostics, especially at preclinical stages, facilitating personalized patient management and targeted clinical trial recruitment based on specific brain endophenotypic expression and prognosis.

The human brain undergoes structural changes during the aging process, but the trajectories of these changes vary markedly among individuals, highlighting the heterogeneity in

---

✉**Correspondence and requests for materials** should be addressed to Christos Davatzikos, Christos.Davatzikos@pennmedicine.upenn.edu.

Author contributions
Study concept and design was by Z.Y. and C.D. Model development was by Z.Y. Data interpretation was by Z.Y., J.W., I.M.N., P.A.L., N.K., S.M.R., H.S. and C.D. Drafting of the manuscript was by Z.Y., J.W., I.M.N. and C.D. Statistical analysis was by Z.Y. Genetic analysis was by J.W. Data collection and processing was by Z.Y., J.W., G.E., S.T.G., R.M., E.M., Y.C., D.S., A.A., P.P., K.W., H.J.G., R.B., S.F., D.T., M.B., Y.A., D.Y., D.S.M., P.L., T.L.S.B., S.R.H., T.R.A., S.R.W., M.K.E., A.B.Z., L.J.L., A.S., M.A.E., C.L.M., P.M., J.F., A.W.T., S.O.B., M.M.C., S.V., S.C.J., J.C.M., M.S.A., K.Y., H.V., L.F., R.N.B., D.A.W., S.M.R. and C.D. Critical revision of the manuscript for important intellectual content was by Z.Y., J.W., G.E., S.T.G., R.M., E.M., Y.C., D.S., A.A., P.P., K.W., H.J.G., R.B., S.F., D.T., M.B., Y.A., D.Y., D.S.M., P.L., T.L.S.B., S.R.H., T.R.A., S.R.W., M.K.E., A.B.Z., L.J.L., A.S., M.A.E., C.L.M., P.M., J.F., A.W.T., S.O.B., M.M.C., S.V., S.C.J., J.C.M., M.S.A., K.Y., H.V., L.F., R.N.B., R.T.S., Y.F., M.H., P.A.L., N.K., D.A.W., S.M.R., H.S., I.M.N. and C.D.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-024-03144-x.

Peer review information *Nature Medicine* thanks Cristina Granziera and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Lorenzo Righetto, in collaboration with the *Nature Medicine* team.

brain aging. The complex interplay of various factors, including genetic and lifestyle factors and diseases, contribute to this heterogeneity, either exacerbating or protecting against age-related neuropathological processes[1]. Moreover, relatively subtle brain changes in specific regions or spatial patterns can emerge early in the preclinical stages of neurodegenerative diseases such as Alzheimer's disease[2,3]. Therefore, it is essential to unravel the heterogeneous neuroanatomical changes associated with aging across a broad spectrum of individuals in a manner that is representative, reproducible and clinically interpretable. Such investigations hold the potential to elucidate the intricate progression of underlying biological, including neuropathological, mechanisms, especially if combined with available direct biomarkers of neuropathology. Additionally, they may yield valuable early markers for diagnosis, prognosis, vulnerability or resilience, guiding the identification of suitable interventions.

Neuroimaging plays a pivotal role in studying the human brain, enabling direct quantification of these changes in vivo[4] and at a large scale. This has enriched our comprehension of how aging and diseases influence brain structure and function. The majority of previous studies have relied on case–control group comparisons, which are not designed to address heterogeneity across individuals and pathologies. Some machine-learning methods, leveraging binary classifications, tried to derive neuroimaging biomarkers of brain aging at the individual level[5,6]; however, these studies still overlook the underlying heterogeneity and derive biomarkers of a typical or averaged pattern of diverse neuroanatomical changes. Various clustering approaches have been deployed to parse the heterogeneity in aging-related neurological diseases from neuroimaging data[7–9]; however, they are usually confounded by numerous variations in brain structure that are not related to aging and neuropathology. Moreover, they aim to cluster each individual into a single subtype, thus overlooking that an individual might have a mixture of underlying pathologies at different stages.

In contrast to previous methods, Surreal-GAN[10], a recent weakly supervised, deep-representation learning method, offers an innovative and general approach applicable to disentangling the heterogeneity of brain aging. By learning multiple transformations from a reference (REF) group (for example, young and healthy individuals) to a target (TAR) group (for example, older adults or patients with a specific clinical phenotype), the model captures heterogeneous brain changes relative to the reference population and effectively distills them down to low-dimensional representation indices. These indices, herein called R-indices, indicate the severity of individualized brain changes along multiple dimensions, potentially reflecting the stage of a mixture of underlying neuropathological and biological processes that induce deviations from the distribution of a reference brain structure.

The contribution of the current study is twofold. First, we substantially extended the Surreal-GAN methodology by introducing a correlation structure among the R-indices in a reduced representation latent space, thereby capturing atrophy manifestations of interactions among multiple (potentially coexisting) underlying neuropathological processes. Second, we applied this methodology to a large and harmonized multi-study, multi-site dataset from the iSTAGING consortium[11], consisting of more than 50,000 participants from 11 neuroimaging studies. As our goal was to capture patterns of brain aging, in our experiments we defined the REF group as participants younger than 50 years old, the age at which a broader spectrum of brain abnormalities[11–13] and diseases affecting the nervous system[14–16] begin to emerge or accelerate. All other individuals above 50 years old, including those with mild cognitive impairment (MCI) or dementia, were grouped as the TAR group. We therefore established a mathematically principled representation of the dominant dimensions of neuroanatomical brain aging in this cohort, and associated these dimensions with cognitive, clinical, lifestyle and genetic measures. Additionally, survival analyses demonstrated that brain change along these dimensions predicted future disease progression and mortality.

Results

Surreal-GAN parses brain aging patterns via latent dimensions

Surreal-GAN[10] utilizes a deep generative model along with a series of effective regularization constraints[10] to learn one-to-many transformations from brain measurements of a REF population, such as a pre-aging or healthy control cohort, to a TAR population, such as aging or disease-related cohorts (Method 1). Surreal-GAN therefore captures dominant dimensions, or patterns, of brain change related to a condition of interest while minimizing confounding variations. The expressions of brain changes along each of these dimensions are called R-indices. The same participant can have non-exclusive coexpression of different patterns, indicating that multiple pathologic processes are potentially active (Fig. 1a).

Improved Surreal-GAN with latent correlations

The fundamental framework of the original Surreal-GAN model inherently encourages independence among the derived R-indices. As demonstrated theoretically (Method 2), this poses a substantial limitation when dealing with correlated ground-truth dimensions or patterns of brain changes. Empirically, we showcased the decline in the model's performance on semi-synthetic data with different levels of simulated associations among ground-truth patterns. This constraint becomes particularly relevant in the context of brain aging due to the common co-occurrence of multiple underlying pathologies and their potential impact on various brain regions.

In the current work, we addressed this limitation by introducing a parameterization of the correlations among the R-indices using a Gaussian copula during the training process (Method 2). The enhanced Surreal-GAN was found to be robust in handling various levels of simulated correlations in semi-synthetic experiments (Supplementary Fig. 1). In real-data experiments, it effectively derived correlated R-indices, which capture the interactions among underlying pathologic mechanisms (Fig. 2c).

Applying Surreal-GAN to unravel brain aging heterogeneity

We defined the training (discovery) set by including a REF group of 1,150 participants of ages 20–49 (43.8 ± 6.6) (pre-aging participants) and a TAR group of 8,992 participants of ages 50–97 (68.2 ± 9.1) including those with MCI or dementia (all older adults) from 11 studies, thus training the Surreal-GAN model to derive dimensional representations that capture the spectrum of brain aging patterns. Subsequently, we applied the resultant model to all 49,482 iSTAGING participants of ages 50–97 years (65.5 ± 7.9) to associate the expression of brain change along each dimension with demographic, clinical, neurocognitive, lifestyle and genetic measures (Fig. 1b). Consistency among independently trained models (Method 4) suggested that the most reproducible R-indices were derived from five dimensions, which were also replicated using an independent training set consisting of 1,000 REF and 4,818 TAR participants (Supplementary Fig. 2). Furthermore, we tested the reproducibility of the five dimensions across sexes by training the models exclusively on male and female participants. The sex-specific R-indices are associated with consistent brain atrophy patterns and demonstrate strong R-indices-correlation (Method 6) values with the original ones (male = 0.886, female = 0.849) (Supplementary Fig. 2). Sex differences are still captured by differences in the expression levels of the five patterns and their relationships with chronological age, as discussed below.

Five R-indices show severity of distinct atrophy patterns

We first investigated neuroanatomical changes related to each index using voxel-based morphometry analyses. R1 exhibited significant associations with subcortical atrophy, mainly concentrated in the caudate and putamen. R2 was characterized by focal medial temporal lobe (MTL) atrophy. R3 indicated the severity of parieto-temporal atrophy, including that in middle temporal gyrus, angular gyrus and middle occipital gyrus. R4 was distinguished by diffuse cortical atrophy in medial and lateral frontal regions, as well as superior parietal and occipital regions. R5 primarily indicated perisylvian atrophy centered around the insular cortex (Fig. 2a).

Additionally, we observed significant positive correlations between white matter hyperintensity (WMH) volumes and R2–R5 (Fig. 2b). Among them, R5 exhibited a much stronger association than other dimensions with and without adjusting for age and sex ($\rho = 0.430$ and $\rho_c = 0.245$, $P < 1 \times 10^{-200}$) Five R-indices displayed significant associations with each other, with the most prominent correlations observed among R3, R4 and R5 (Pearson's r values 0.39–0.46; Fig. 2c).

Association with demographics and education attainment

All R-indices showed significant correlations with chronological age, with the strongest correlations observed between R5 and age (Fig. 2d). In contrast, R1 and R2 showed relatively lower correlations. After adjusting for age, significant differences emerged between male and female groups in all five R-indices, with the most notable differences observed in R3 ($d = 0.171$) and R5 ($d = 0.178$). Male participants showed greater brain atrophy than females in all dimensions, as reflected by larger R-indices. Additionally, tested through the Fisher r-to-z transformation, we observed small but significant differences between male and female groups in correlations between age and R1 ($P = 1.6 \times 10^{-4}$), R3 ($P = 2.7 \times 10^{-5}$) and R5 ($P = 7.6 \times 10^{-4}$), suggesting potential influences from sex-related factors. Education attainment has been an important factor in understanding brain function; however, we observed significant associations only between R3 and education attainment with a small effect size ($P = 0.026$). Therefore, we chose not to correct for education attainment in subsequent analyses.

Association with chronic diseases

To investigate the relationship between R-indices and chronic disease risk, UK Biobank (UKBB) individuals with a lifetime diagnosis of a chronic disease were grouped into 14 disease categories: MCI/dementia, stroke, multiple sclerosis, hypertensive diseases, diabetes, depression, bipolar disorder, schizophrenia, Parkinson's disease, chronic obstructive pulmonary disease (COPD), osteoarthritis, chronic kidney disease (CKD), osteoporosis and ischemic heart disease. An additional 2,550 participants from six other studies were included to enrich the MCI/dementia cohort and an additional (age-compatible) 71 participants from the PHENOM consortium[17] (Psychosis Heterogeneity Evaluated via Dimensional Neuroimaging) were included to enrich the schizophrenia cohort. UKBB participants without any of the 14 diseases were categorized as a healthy control (HC) group. To test differences in R-indices between the control and diseased groups, we conducted multiple linear regression analyses, adjusting for age and sex (Fig. 3a).

The MCI/dementia cohort exhibited significantly advanced R-indices along all five dimensions, with the most prominent effects observed in R2 (Cohen's $d = 0.869$, $P = 4.1 \times 10^{-131}$), R3 ($d = 0.849$, $P = 1.6 \times 10^{-125}$) and R5 ($d = 0.673$, $P = 5.7 \times 10^{-81}$). Patients with multiple sclerosis showed advanced aging along R3 ($d = 0.498$, $P = 6.7 \times 10^{-8}$), R4 ($d = 0.515$, $P = 2.4 \times 10^{-8}$) and R5 ($d = 0.513$, $P = 2.8 \times 10^{-8}$). Additionally, participants with schizophrenia (R3: $d = 0.507$, $P = 2.5 \times 10^{-8}$; R5: $d = 0.725$, $P = 2.0 \times 10^{-15}$) and Parkinson's disease (R3: $d = 0.481$, $P = 9.6 \times 10^{-6}$; R5: $d = 0.370$, $P = 6.6 \times 10^{-4}$) demonstrated more brain atrophy along the R3 and R5 dimensions, with schizophrenia participants also showing higher R2 values ($d = 0.566$, $P = 5.3 \times 10^{-10}$). Furthermore, the R5 dimension was associated with neuropsychiatric diseases, as well as a group of chronic diseases related to other organ systems, including the respiratory, renal, metabolic and cardiovascular systems. Among hypertension, diabetes, CKD and ischemic heart disease, common comorbidities, additional correction of the other three did not lead to substantial changes in associations with hypertension and diabetes; however, in R5, differences became statistically insignificant between HC and patients with CKD ($d = 0.101$, $P = 0.021$) or ischemic heart disease ($d = 0.056$, $P = 0.012$), suggesting that hypertension and diabetes are the main factors affecting brain changes along the R5 dimension in people with these conditions. Detailed results between R-indices and all diseases can be found in Supplementary Data 1.

Baseline R-indices predict disease progression and mortality

In addition to exploring baseline associations with chronic diseases, we examined the prognostic potential of R-indices for future progression from cognitively normal (CN) to MCI and from MCI to dementia using 2,700 participants with longitudinal data from seven studies. The Cox proportional hazard model was utilized to test the associations between R-indices and the risk of progression, while adjusting for age and sex. R2 ($P = 2.7 \times 10^{-8}$, hazard ratio (HR) 4.50, 95% confidence interval (CI) 2.65–7.64), R4 ($P = 4.5 \times 10^{-3}$, HR 2.05, 95% CI 1.25–3.37) and R5 ($P = 1.3 \times 10^{-4}$, HR 2.93, 95% CI 1.69–5.09) were significantly associated with the risk of CN to MCI conversion, with R2 demonstrating the strongest prognostic indicator (Fig. 3b). R2 ($P = 1.1 \times 10^{-26}$, HR 6.37, 95% CI 4.54–8.94), R3 ($P = 2.8 \times 10^{-19}$, HR 7.97, 95% CI 5.07–12.54) and R5 ($P =$

2.2×10^{-8}, HR 3.70, 95% CI 2.34–5.86) were significantly associated with the risk of progression from MCI to dementia, with R2 and R3 being the leading prognostic indices (Fig. 3b). In cross-validation on participants over 60 years old, combining R2 (Concordance Index (C-index) 0.768 ± 0.031) and R5 (C-index 0.773 ± 0.033) as features in addition to age and sex (C-index 0.755 ± 0.033) improved the prediction for the risk of CN to MCI progression. Similarly, for predicting progression from MCI to dementia, the addition of R3 (C-index 0.646 ± 0.035), R2 (C-index 0.704 ± 0.030) and R5 (C-index 0.705 ± 0.030) significantly enhanced predictive performance (C-index 0.550 ± 0.034) (Fig. 3c). Furthermore, in Fig. 3d, we showcased the longitudinal progression paths of R2, R3 and R5 for eight representative participants who experienced disease conversions. In alignment with the baseline risk associations, these representative participants exhibited rapid increases in R2 before transitioning from CN to MCI, and rapid increases in R2 and R3 before progressing from MCI to dementia. Significant increases in R5 were also observed among some participants following their diagnoses of dementia.

Using similar approaches, we then examined the risk of mortality using UKBB participants (Fig. 3e). R3 ($P = 1.8 \times 10^{-3}$, HR 1.95, 95% CI 1.28–2.98), R4 ($P = 3.2 \times 10^{-3}$, HR 1.89, 95% CI 1.24–2.87) and R5 ($P = 1.3 \times 10^{-6}$, HR 2.98, 95% CI 1.91–4.64) showed significant associations with the risk of mortality, with R5 being the most significant prognostic indicator. Among participants over 60 years old, the combination of R5 (C-index 0.672 ± 0.029), R3 and R4 (C-index 0.674 ± 0.029) as features, along with age and sex (C-index 0.661 ± 0.027), improved the prediction for the risk of mortality, while the subsequent addition of other R-indices resulted in slightly decreasing predictive performances.

Associations with clinical variables

We further explored associations between the five dimensions and cognition, as well as cerebrospinal fluid (CSF)/plasma biomarkers. Using partial correlations, we first tested associations of R-indices with four Alzheimer's Disease Neuroimaging Initiative (ADNI) composite scores within the ADNI cohort ($n = 2{,}214$), along with four other cognitive scores among participants from multiple studies ($n = 6{,}280$–$30{,}444$; Fig. 4a). We observed consistent and significant associations among R2, R3 and R5 and all cognitive scores, although the pattern of associations differed between the cognitive scores. Specifically, R2 exhibited a particularly pronounced correlation with memory performance ($\rho = 0.462$, $P = 1.8 \times 10^{-117}$, measured by the memory (MEM) composite of ADNI (ADNI-MEM), while R3 demonstrated similar associations with executive function ($\rho = 0.349$, $P = 9.8 \times 10^{-65}$ versus the executive function (EF) composite (ADNI-EF) and memory ($\rho = 0.343$, $P = 3.2 \times 10^{-62}$ versus ADNI-MEM). Detailed correlations and P values can be found in Supplementary Data 2.

Among the 232 CSF/plasma biomarkers collected in the ADNI study, we identified 15 significant associations between R-indices and biomarkers (Fig. 4b). Among them, R2 and R3 revealed significant positive associations with CSF-pTau181 ($P = 1.6 \times 10^{-12}$ versus R2; $P = 1.3 \times 10^{-14}$ versus R3) and negative associations with CSF-Aβ42 ($P = 2.6 \times 10^{-23}$ versus R2; $P = 4.4 \times 10^{-23}$ versus R3), two hallmarks of Alzheimer's disease[18]. Binary amyloid and tau positivity also showed significant associations with R2 ($\rho = 0.290$, $P = 7.0$

$\times 10^{-28}$ versus amyloid^{+}; $\rho = 0.173$, $P = 1.4 \times 10^{-10}$ versus tau^{+}) and R3 ($\rho = 0.258$, $P = 4.1 \times 10^{-22}$ versus amyloid+; $\rho = 0.110$, $P = 4.3 \times 10^{-5}$ versus tau^{+}). R5 was associated with a group of other biomarkers, including chromogranin-A ($P = 1.3 \times 10^{-11}$), tissue factor ($P = 2.6 \times 10^{-7}$), AXL receptor tyrosine kinase (AXL) ($P = 3.5 \times 10^{-5}$), angiotensin-converting enzyme ($P = 5.4 \times 10^{-5}$), cystatin-C ($P = 1.1 \times 10^{-5}$) and interleukin(IL)-6 receptor ($P = 2.4 \times 10^{-4}$), which potentially reflected the underlying hemostatic and inflammatory mechanisms[18–20]. Detailed correlations and P values can be found in Supplementary Data 3.

Association with environmental and lifestyle factors

We examined the influence of lifestyle and environmental factors on variations along the five dimensions. Using partial correlations, we assessed the associations between the R-indices and 120 variables from the UKBB study, adjusting for age and sex (Fig. 4c). In relation to all five atrophy dimensions, we found that alcohol intake has a statistically significant association with brain atrophy. Furthermore, smoking status primarily was associated with the expressions of R3 to R5. Notably, R4 and R5 were the two dimensions most associated with these two types of lifestyle factors. Moreover, the R3 dimension revealed significant associations pregnancy and social-recreational activities; the R4 dimension showed additional associations with various dietary habits; and the R5 dimension was exclusively associated with long-term illness, emotional factors, sleep and environmental factors. Detailed correlations and P values can be found in Supplementary Data 4.

The five R-indices were associated with genetic variants

Genome-Wide Association Study (GWAS) (Method 15a) identified 16, 17, 13, 9 and 18 genomic loci significantly associated with R1–R5, respectively (Fig. 5a, Supplementary Fig. 3 and Supplementary Data 5). Among them, 38 loci were never associated with any clinical traits in the EMBL-EBI GWAS Catalog[21], including 11, 7, 6, 5 and 11 loci for the five R-indices, respectively.

Using the GWAS catalog, we performed a phenome-wide association query (Method 15b) of the genetic variants previously identified in our GWAS. Specifically, genetic variants in our GWAS were previously associated with a total of 3,895 clinical traits related to multiple organ systems and lifestyle factors. As expected, the traits associated with the genetic variants of all five R-indices primarily involve different brain region volumes[22] and the microstructure characteristics of white matter tracts as measured by diffusion tensor imaging[23]. Additionally, these genomic loci were also enriched in many traits related to other organ systems (Fig. 5b). For instance, R1 and R5 loci were enriched in traits related to body shape and cardiovascular system. R3 and R5 loci were linked to traits related to psychiatry and psychology, including schizophrenia, depression and worry feelings. Also, R5 loci were specifically associated with WMH-related traits. Notably, through gene-set enrichment analyses (Method 15c), we found that the R1 dimension was enriched in the biological pathway of response to cortisol ($P = 1.1 \times 10^{-6}$), which is related to the mediation of stress.

Discussion

Human brain aging is affected heterogeneously by a complex interplay of genetic[24–27], lifestyle[28,29] and pathological factors[30,31], leading to diverse neuroanatomical changes. Previous research can be categorized into disease-focused studies and population studies. Disease-focused investigations often examine brain atrophy patterns associated with specific pathologies, such as Alzheimer's disease[8,32,33] and neuropsychiatric diseases[34,35], focusing on the nuanced variations within the strictly defined cohorts. Population studies analyze a wider range of patterns using general populations, sometimes without confirmed specific pathologies or within contexts of low disease prevalence[36–40]. Our research is an evolution of population studies, employing a recently developed deep-representation-learning method, Surreal-GAN, to find data-driven patterns of change, on a diverse pool of harmonized magnetic resonance imaging (MRI) data from nearly 50,000 older adults. Using this method, we delineate the predominant, reproducible patterns of brain atrophy observed in the human aging process. These patterns, presumed to be influenced by interactions between various diseases and intrinsic factors, were further examined through post hoc analyses to illuminate their genetic, clinical, cognitive and lifestyle correlates.

Furthermore, we extended the Surreal-GAN model by adding a correlation structure in its latent space, which allowed us to better disentangle co-occurring and partially overlapping patterns of brain atrophy. The improved method identified five reproducible dimensions of neuroanatomical changes termed R-indices: R1, subcortical atrophy; R2, MTL atrophy; R3, parieto-temporal atrophy; R4, diffuse cortical atrophy; and R5, perisylvian atrophy. Critically, our approach allows for evaluation of individualized levels of expression along these five dimensions, thereby offering additional tools for personalized patient management and clinical trial stratification. While we did observe distinct and overlapping patterns of associations of R-indices with diseases and risk factors, we also found correlations among the expression of the five dimensions, indicating coexpression of corresponding brain atrophy patterns at varying degrees. Together, these underscore the interconnection and co-occurrence of underlying biological or neuropathological mechanisms[30,41] and suggest that the R-indices potentially measure the impact of multiple co-pathologies on the brain at an individual level.

The dimensional system illuminates the intricate relationships between pathological factors and variations in brain aging. Multiple sclerosis shows correlation with dimensions R3 to R5, while sparing deep structures and medial temporal lobes, aligning with previous findings of diffuse cortical thinning in this disease[42,43], attributed to pathological lesions disrupting structural pathways and the global brain networks[44]. R3 and R5 are also associated with schizophrenia and Parkinson's disease, aligning with established patterns of atrophy[45–47]. Specifically, in schizophrenia, middle temporal gyrus atrophy, contributing to R3, is linked to auditory–verbal hallucinations, a prominent symptom of this disorder[48]. R5 exhibits broad associations with various systemic diseases, including neuropsychiatric disorders, cardiovascular diseases and immune health factors, and shows the strongest associations with WMH volumes. These findings might be partially explained by the role of insular cortex in autonomic regulation[49–51] and emotion processing[52,53]. Recent studies have increasingly recognized the role of the insula in depression, bipolar disorder and schizophrenia, specifically related to disordered interoceptive function[54]. Depression, on the other hand, has been linked to white matter injuries that may have a causal role in its etiology[55]. Furthermore, the co-occurrence of distinct diseases in the same dimensions can provide insights into their shared symptoms and the increased mutual risk. For example, schizophrenia, Parkinson's disease and dementia, presenting overlap in associated dimensions, exhibit common symptoms such as cognitive decline and hallucinations, with schizophrenia and Parkinson's disease associated with a higher risk of dementia[56,57].

Projecting patient cohorts with specific diseases onto our dimensional system offered insights into disease heterogeneity. Various chronic diseases, including MCI/dementia, multiple sclerosis and Parkinson's disease, exhibited elevated expression of multiple dimensions. Individualized differences along these associated dimensions likely mirror their distinct phenotypic and pathological variations. For instance, AD[58], a prevalent neurodegenerative condition among the elderly, presents considerable heterogeneity[32,33,59]. Among the five dimensions, R2, R3 and R5 are strongly correlated with MCI/dementia and display differential associations with mainly the AD-typical characteristics, including cognitive decline and abnormal amyloid and tau deposition[60]. R2 displays stronger associations with memory impairment, whereas R3 is more closely associated with executive dysfunction, features that are differentially expressed in patients with AD. In contrast to R2 and R3, R5 reveals weaker associations with amyloid and tau, but stronger associations with vascular pathology and CSF biomarkers of inflammation, which is an additional feature of AD[18–20], perhaps providing a measure of combinatorial effects of co-pathology on atrophy in AD[61]. Clustering based on these concise R-indices may allow identification of more homogeneous subgroups for clinical trial recruitment or personalized treatment strategies.

Brain changes occurring along disease-associated dimensions may manifest during preclinical stages and provide prognostic values for future disease conversion and mortality. R2, R3 and R5 indicate the risk of clinical progression along the AD continuum. R2 and R3 stand out as key predictors for the clinical conversion from CN to MCI and MCI to dementia, respectively. These results are consistent with the known early involvement of the hippocampus and MTL (R2) in AD and subsequent spread of the pathology to posterior parietal regions (R3)[62]. Concerning mortality risk, chronological age and male sex emerge as the strongest risk factors. Controlling for them, baseline R-indices, particularly R5, retain significant prognostic values. These findings further underscore not only the clinical applicability of R-indices in disease prognosis but also the significance of uncovering interventions targeting factors that are associated with these dimensions to mitigate risk.

Our analyses of lifestyle and environmental factors elucidated additional correlates of the R-indices, thus suggesting potential interventions to target specific dimensions. Smoking and alcohol consumption, two important risk factors across many diseases, are negatively associated with cortical atrophy, primarily mapping to the R4 and R5 dimensions, with lesser effects on R3. These findings align with previous studies that revealed global atrophy with increased alcohol consumption[63,64] and cigarette smoking[65,66]. In addition, daily dietary habits correlate with R4 and R5, having either negative (tea and cereal) or positive (cheese, coffee and salt) associations, in line with previous observations[67]. Directly managing lifestyle or investigating underlying mechanisms might yield feasible interventions, although further studies are needed to understand causal relationships. Expression of any of the five intermediate phenotypes identified herein can serve as an indicator of active involvement of respective genetic and lifestyle risk factors, thereby prompting more aggressive patient management as well as recruitment to respective clinical trials.

Beyond daily life factors, certain life experiences also have associations with brain aging across dimensions. The association of R1 and R3 with pregnancy-related factors partially explain their slightly larger deviation from chronological age in females. Notably, childbirth has been shown to exert a long-term influence on women's brain age in late life[68], possibly contributing to a 'young-looking' brain presumed to be partially related to fluctuations in hormonal or inflammatory mechanisms. Our study advances understanding by mapping these effects to specific dimensions of brain changes in addition to the overall brain age. In CN participants, R2 was negatively correlated with birth weight, resonating with previous findings of reduced hippocampal volumes and learning difficulties among preterm-born children with very low birth weights[69].

Genetic variants influencing brain aging heterogeneity provide additional insights for intervention strategies, particularly in drug development. Our genetic analyses have revealed 73 single-nucleotide polymorphisms (SNPs) associated with the five dimensions, including 38 SNPs without associated traits in the GWAS catalog. The previously identified SNPs correlate with various clinical traits that validate our findings. For instance, R3 loci are linked to schizophrenia-related traits, whereas R5 loci show associations with cardiovascular and neuropsychiatric conditions, along with white matter hyperintensities. Additionally, the R1 dimension displays a correlation with the gene set linked to the response to cortisol. This suggests a potential stress-related impact on morphological changes in the striatum, the region connected to R1 and demonstrated to be highly influenced by stress[70,71]. The identification of these SNPs and associated genes might inform drug discovery or repurposing efforts for interventions targeting these dimensions.

The five dimensions derived herein are limited by the resolution and detail offered by MRI, namely patterns of regional atrophy of gray matter and white matter, measures of small vessel ischemic disease and expansion of CSF spaces. As such, they do not directly measure underlying neuropathologic processes that lead to these brain atrophy; however, they do allow us to directly assess the personalized impact of specific risk factors on brain changes through more precise and standardized measurements. They can therefore offer opportunities for personalized patient management and clinical trial recruitment (Fig. 6). In particular, the dimensions represented by R-indices enable the provision of more personalized therapeutic plans and lifestyle recommendations tailored to individual expression levels. Moreover, beyond facilitating targeted patient recruitment, the five R-indices also contribute to more optimized clinical trial design from several other perspectives. For instance, tracking changes in specific R-indices can contribute to the assessment of trial effectiveness and boost its power[72]. Furthermore, a combination of R-indices and other clinical variables can be employed to establish stratified trial benchmarks due to their indication of individualized disease progression speed at the trial's baseline.

While our study has included a broad and diverse population for training and analysis, it is essential to acknowledge certain limitations. One constraint is the underrepresentation of uncommon pathologies, limiting the representation of all neuropathology by the current five patterns. Additionally, the restricted sample size for certain diseases, such as Parkinson's disease, may further affect the power of our analyses on correlating R-indices with pathology. Furthermore, our choice of age threshold was guided by both known aging trends and data availability, given the relatively limited number of individuals below 40 years old. We recognize that this approach may obscure certain preclinical changes within the reference group, potentially reducing the power of our study.

Despite these limitations, the enhanced Surreal-GAN methodology demonstrates broad applicability and substantial potential for pattern discovery, which allows for optimized adaptation tailored to various research questions, accommodating differences in data availability and curation across different study designs. Moreover, our current five-dimensional coordinate system can serve as a foundation for future brain aging research, by offering higher specificity in investigating relationships with brain functional networks via functional MRI and underlying biological mechanisms through genomic, transcriptomic and proteomic data. Additionally, it sets the stage for an expanded future system further enriched by neuroanatomical heterogeneity found in patient populations not amply represented in our current sample, including frontotemporal dementia, dementia with Lewy bodies and other neurodegenerative diseases. The continuous expansion of this neuroanatomical coordinate system, as well as its free-and-easy web access through the cloud, are among the main goals of our neuroimaging brain chart (neuroimagingchart.com).

In summary, this study characterized the neuroanatomical heterogeneity related to the aging process by offering a five-dimensional representation system with five distinct brain signatures. As our consortium analyzes additional imaging data, including diffusion and functional MRI and tau and amyloid positron emission tomography, it will continue to be extended and enriched. Currently, this dimensional system offers a means for dissecting the heterogeneity of brain atrophy, as captured by structural MRI, and to further understand its relationships to demographic, pathological and lifestyle factors, as well as genetic variants. Moreover, it may contribute to personalized diagnostics and patient management, as well as to increased precision and effectiveness of clinical trials.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-024-03144-x.

Methods

Method 1: Surreal-GAN model

The Surreal-GAN method[10] is a weakly supervised deep-representation-learning method for disentangling disease heterogeneity from neuroimaging data. Its key advantage lies in the ability to discern spatial and temporal (disease severity) variations solely from baseline data, thereby deriving low-dimensional R-indices that directly indicate the severity of distinct patterns of neuroanatomical changes. To capture phenotypic changes due to disease effects, Surreal-GAN learns multiple transformations from a REF group (for example, pre-aging or health control) to a TAR group (for example, aging or patient)) through the generative adversarial network (GAN). Specifically, the method learns a function f to transform the REF data $\mathbf{x}$ into generated TAR data $\mathbf{y}' = \mathrm{f}(\mathbf{x}, \mathbf{z})$, where $\mathbf{z}$ is a latent variable indicating the transformation directions. As is common in GAN-related methods[73], an adversarial discriminator D is introduced to distinguish between real TAR data $\mathbf{y}$ and synthesized TAR data $\mathbf{y}'$, thereby ensuring that the generated image data are indistinguishable from real patient data.

Beyond that, an inverse mapping, g, is introduced to re-estimate the latent variables $\mathbf{z}$ from the generated data $\mathbf{y}'$ to ensure that the latent variables capture distinct and recognizable brain signatures. Multiple other regularizations were employed to further encourage the transformation function f to approximate the disease or aging effect, while boosting the positive association of different dimensions of the variable ($\mathbf{z}$) with atrophy severity in distinct brain regions.

In the stage of model inference following training, the inverse function, g, is utilized to derive the latent variables (referred to as R-indices) for real TAR data after the training process. Through GAN and regularizations during the training process, the learned transformation function f was considered a good approximation of the underlying pathological process, denoted by function h, such that $\mathrm{f}(\mathbf{x}, \mathbf{z}) \approx \mathrm{h}(\mathbf{x}, \sigma(\mathbf{z}))$, where $\sigma \in \Omega$ and Ω is a class of permutation functions that change the order of elements in the latent variables $\mathbf{z}$. As the orders of indices in the latent variables are unimportant and we can always reorder them to find the best matching, we rewrite the equation as $\mathrm{f}(\mathbf{x}, \mathbf{z}) \approx \mathrm{h}(\mathbf{x}, \mathbf{z})$ without loss of generality. For any real TAR data, $\bar{\mathbf{y}} = \mathrm{h}(\bar{\mathbf{x}}, \mathbf{r}) \sim \mathrm{p}_{\mathrm{tar}}(\mathbf{y})$, we can estimate its ground-truth R-indices, represented by $\mathbf{r}$, through $\mathrm{g}(\bar{\mathbf{y}}) = \mathrm{g}(\mathrm{h}(\bar{\mathbf{x}}, \mathbf{r})) \approx \mathrm{g}(\mathrm{f}(\bar{\mathbf{x}}, \mathbf{r})) \approx \mathbf{r}$. More methodological details can be found in Yang et al.[10].

Method 2: improvement of Surreal-GAN to enable correlations in its latent space

The original Surreal-GAN model[10] derives independent R-indices, which limits its ability to characterize atrophy patterns driven by associated underlying pathologies. Specifically, In most GAN-based models[73,74], the latent variables or noise variables are sampled from fixed distributions (for example, Gaussian or uniform distribution) predefined before the training process. The fixed latent distributions mostly do not affect the models' performances in generating realistic data; however, they are problematic if we use an inverse function to re-estimate the latent variables of the data.

As introduced in Method 1, we assume the inverse consistency, $\mathrm{g}(\mathrm{f}(\mathbf{x}, \mathbf{z})) \approx \mathbf{z}$ and equality in distributions, $\mathrm{p}_{\mathrm{syn}}(\mathrm{f}(\mathbf{x}, \mathbf{z})) \approx \mathrm{p}_{\mathrm{tar}}(\mathbf{y})$, after the training process. We can further derive that $\mathrm{p}(\mathrm{g}(\mathbf{y})) \approx \mathrm{p}(\mathrm{g}(\mathrm{f}(\mathbf{x}, \mathbf{z}))) \approx \mathrm{p}(\mathbf{z})$, where $\mathrm{p}(\mathrm{g}(\mathbf{y}))$ is the distribution of the derived TAR participants' R-indices. Therefore, if we sample $\mathbf{z}$ from a standard multivariate uniform distribution as in the original Surreal-GAN, the covariance of the derived R-indices will be the identity matrix, which leads to bias and decreased model performances when the ground-truth R-indices are correlated with each other (Supplementary Fig. 1).

To resolve this issue, we first constructed a parametrized latent distribution for **z** using Gaussian copula, denoted as $p_{\theta_z}(\mathbf{z}) = C_{\theta_z}^{\mathrm{Gauss}}(\mathbf{z})$, where the learnable parameters θ_z govern the correlations among dimensions:

$$p_{\theta_z}(\mathbf{z}) = C_{\theta_z}^{\mathrm{Gauss}}(z) = \Phi_{\theta_z}\left(\Phi^{-1}(z_1),\ \Phi^{-1}(z_2),\ \dots,\ \Phi^{-1}(z_M)\right) \tag{1}$$

where Φ^{-1} is the inverse cumulative distribution function of a standard normal and Φ_{θ_z} is the joint cumulative distribution function of a multivariate normal distribution with the mean vector zero and the covariance matrix equal to a correlation matrix $\theta_\mathbf{z}$. The resulting distribution, $p_{\theta_z}(\mathbf{z})$, has covariance equals $\theta_\mathbf{z}$ and has the marginal distribution of each dimension to be uniform, $U[0, 1]$. To optimize this parametrized latent distribution to approximate the ground-truth correlations, we introduced modifications to the original GAN loss function,

$$L_{\mathrm{GAN}}(\theta_D,\ \theta_f,\ \theta_z) = E_{y \sim p_{\mathrm{tar}}(y)}[\log(D(\mathbf{y}))] + E_{y' \sim p_{\mathrm{syn}}(y')}[1 - \log(D(\mathbf{y}'))] \tag{2}$$

$$= E_{z \sim p_{\theta_z}(z),\, y \sim p_{\mathrm{tar}}(y)}[\log(D(\mathbf{y}))] + E_{z \sim p_{\theta_z}(z),\, x \sim p_{\mathrm{ref}}(x)}[1 - \log(D(f(\mathbf{x},\ \mathbf{z})))] \tag{3}$$

$$\begin{aligned} = & \int_{R^M} p_{\theta_z}(\mathbf{z}) E_{y \sim p_{\mathrm{tar}}(y)}[\log(D(\mathbf{y}))] d^M\mathbf{z} \\ + & \int_{R^M} p_{\theta_z}(\mathbf{z}) E_{x \sim p_{\mathrm{ref}}(x)}[1 - \log(D(f(\mathbf{x},\ \mathbf{z})))] d^M\mathbf{z} \end{aligned} \tag{4}$$

$$\begin{aligned} = & \int_{R^M} p_U(\mathbf{z}) p_{\theta_z}(\mathbf{z}) E_{y \sim p_{\mathrm{tar}}(y)}[\log(D(\mathbf{y}))] d^M\mathbf{z} \\ + & \int_{R^M} p_U(\mathbf{z}) p_{\theta_z}(\mathbf{z}) E_{x \sim p_{\mathrm{ref}}(x)}[1 - \log(D(f(\mathbf{x},\ \mathbf{z})))] d^M\mathbf{z} \end{aligned} \tag{5}$$

$$\begin{aligned} = & E_{y \sim p_{\mathrm{tar}}(y),\, z \sim p_U(z)}\left[p_{\theta_z}(\mathbf{z}) \log(D(\mathbf{y}))\right] \\ + & E_{z \sim p_U(z),\, x \sim p_{\mathrm{ref}}(x)}\left[p_{\theta_z}(\mathbf{z})(1 - \log(D(f(\mathbf{x},\ \mathbf{z}))))\right] \end{aligned} \tag{6}$$

The updated loss function enables us to sample $\mathbf{z}$ from a multivariate uniform distribution, $\mathrm{p}_{\mathrm{U}}(\mathbf{z}) = \mathrm{U}[0,1]^{M}$, but penalizes the losses with their probability under the distribution $\mathrm{p}_{\theta_z}(\mathbf{z})$. In the training procedure, we sample one value of $\mathbf{z}$ for each $\mathbf{x}$ per batch instead of taking the expectation over all possible values of $\mathbf{z}$. Therefore, to penalize both terms equally, the same modification is also applied to the first term of the equation (6), in which $\mathbf{z}$ was not originally included.

Additionally, to prevent p_{θ_z} from converging to an extreme distribution (for example, two latent variables become completely correlated), we added a regularization term that controls the distance between $\mathrm{p}_{\theta_z}(\mathbf{z})$ and $\mathrm{p}_{\mathrm{U}}(z)$. Therefore, the final modified GAN loss function equals:

$$\begin{aligned} \mathrm{L}_{\mathrm{GAN}}(\theta_{\mathrm{D}},\ \theta_{\mathrm{f}},\ \theta_{\mathrm{z}_2}) \ = \ & \mathrm{E}_{\mathrm{y} \sim \mathrm{p}_{\mathrm{tar}}(\mathrm{y}),\, \mathrm{z} \sim \mathrm{p}_{\mathrm{U}}(\mathrm{z})}[\mathrm{p}_{\theta_{\mathrm{z}}}(\mathbf{z})\log(\mathrm{D}(\mathbf{y}))] \\ & + \ \mathrm{E}_{\mathrm{z} \sim \mathrm{p}_{\mathrm{U}}(\mathrm{z}),\, \mathrm{x} \sim \mathrm{p}_{\mathrm{ref}}(\mathrm{x})}\left[\mathrm{p}_{\theta_{\mathrm{z}}}(\mathbf{z})(1 - \log(\mathrm{D}(\mathrm{f}(\mathbf{x},\ \mathbf{z}))))\right] \\ & + \ \alpha \mathrm{D}_{\mathrm{KL}}\left(\mathrm{p}_{\mathrm{U}}(\mathbf{z}) \middle| \mathrm{p}_{\theta_{\mathrm{z}}}(\mathbf{z})\right) \end{aligned} \tag{7}$$

This improvement enables the identification of associated R-indices, resulting in a notable boost in model's performances and enhancing the method's applicability (see Supplementary Fig. 1 for details).

Method 3: study population

The MRI (Method 9) and clinical (Method 10 and 11) data used in this study were consolidated and harmonized by the Imaging-Based Coordinate System for Aging and Neurodegenerative Diseases (iSTAGING) study. The iSTAGING study comprises data acquired via various imaging protocols, scanners, data modalities and pathologies, including more than 50,000 participants from more than 13 studies on three continents and encompassing a wide range of ages (22 to 90 years). Specifically, the current study used data from 52,319 participants from the ADNI[75], UKBB[76], the Baltimore Longitudinal Study of Aging (BLSA)[77,78], the Australian Imaging, Biomarker, and Lifestyle study of aging (AIBL)[79], the Biomarkers of Cognitive Decline Among Normal Individuals in the John Hopkins University study (BIOCARD)[80], the Open Access Series of Imaging Studies (OASIS)[81], the University of Pennsylvania Memory Center cohort (PENN), the Wisconsin Registry for Alzheimer's Prevention (WRAP) studies[82], the Coronary Artery Risk Development in Young Adults (CARDIA)[83], Study of Health in Pomerania (SHIP)[84] and the Women's Health Initiative Memory Study (WHIMS)[85]. Among them, longitudinal data were available for 6,576 participants. Detailed demographics and sample sizes from each study are detailed in Table 1. MRI details of all studies are provided in Supplementary Data 6.The aging participants analyzed in this study were predominantly of white ethnicity ($n = 44,539$), constituting 95.1% of the total participants with reported race information. Race information was provided by all studies except SHIP. Detailed race distributions across different studies were provided in Supplementary Data 7, with race classification criteria introduced in Supplementary Data 8. Participants provided written informed consent to the corresponding studies. The protocols of this study were approved by the University of

Pennsylvania institutional review board. The ADNI database, launched in 2003 and led by Principal Investigator M. W. Weiner, aims to integrate various assessments, including MRI, positron emission tomography and clinical evaluations, to monitor the progression of MCI and early AD. Updates are provided at https://adni.loni.usc.edu/.

Method 4: R-indices derivation

For training the Surreal-GAN model, we included baseline data from 1,150 participants below 50 years old as the REF group and that of 8,992 participants over 50 years old as the TAR group including those with MCI or dementia. We noticed the imbalance in sizes between the REF and TAR groups; however, in this process, we prioritize the size and coverage of each group individually so that the selected samples better represent data distributions. We aimed to include as much data as possible in each group while maintaining a more even distribution across studies to prevent dominance from any single study. Thus, a maximum of 300 and 1,000 individuals from each study were included in the REF and TAR groups. The larger sample size of the TAR group could also better cover the distribution of the aging population with a larger variance due to heterogeneity. All individuals in the REF and TAR groups were first residualized to rule out the sex and intracranial volume (ICV) effects estimated in the REF group using linear regression. Then, adjusted features were standardized with respect to the REF group. Without ground truth, we selected both the optimal number of dimensions, M (2–7) and two important hyperparameters[10], γ and λ, by measuring agreements (Method 6) among repetitively trained models following the selection procedure introduced in the original Surreal-GAN paper[10]. In unsupervised learning (for example, representation learning or clustering), reproducibility is paramount to establishing the reliability of derived representations and subsequent analyses. Also, as demonstrated in Supplementary Table 1, the agreements among models can indicate the underlying representation accuracies. For each of combination of hyperparameters, we repetitively trained the model 50 times and determined the optimal hyperparameters leading to the highest agreements among the 50 models ($M = 5, \gamma = 0.1, \lambda = 0.8$). Among the 50 corresponding models, the one having the highest mean pair-wise agreement with the other models was used to derive the five R-indices for all 49,482 elderly participants.

Method 5: Surreal-GAN implementation details

We set five robust hyperparameters, $\alpha = 0.02, \kappa = 80, \zeta = 80, \mu = 500$ and $\eta = 6$, to their default values, which minimally impact model performance[10]. For optimal training epoch selection, the model underwent $1,500,000 \times \frac{\text{batch size}}{\text{TAR sample size}}$ epochs, with checkpoints saved every $45,000 \times \frac{\text{batch size}}{\text{TAR sample size}}$ epochs when reconstruction loss and monotonicity loss were below 3×10^{-3} and 6×10^{-4}, respectively. Same to other hyperparameters, agreements among repeatedly trained models once again guided the determination of the optimal epoch.

Regarding the optimization procedure, the ADAM optimizer was used with a learning rate (lr) 1.6×10^{-4} for discriminator, 8×10^{-4} for transformation function f and inverse function g, and 2.7×10^{-5} for the latent distribution, $p_{\theta_z}(z)$. β_1 and β_2 are set to be 0.5

and 0.999, respectively. Moreover, for all experiments, the batch size was set to 300. We parametrize the latent distribution (the Gaussian copula) using the lower-triangular factor of the correlation matrix, given its special properties. Other model structures and training algorithms were the same as in the original Surreal-GAN[10].

Method 6: agreement metric

For hyperparameter selection and result reproducibility evaluation, we constructed a metric designed to quantify the level of agreement between two sets of R-indices, $\mathbf{r}^1$ and $\mathbf{r}^2$. To assess their concordance, we computed two distinct correlations, as outlined below:

1. Dimension-correlation is defined as the average of M Pearson's correlations for all dimensions: $\frac{1}{M}\left(\sum_{i=1}^{M}\rho\left(\mathrm{r}_i^1, \mathrm{r}_i^2\right)\right)$
2. Difference-correlation is defined as the average of $M(M-1)/2$ Pearson's correlations for all pairs of dimensions: $\frac{2}{M(M-1)}\left(\sum_{i=1}^{M}\sum_{j=i+1}^{M}\rho\left(\mathrm{r}_i^1 - \mathrm{r}_j^1, \mathrm{r}_i^2 - \mathrm{r}_j^2\right)\right)$

For the derivation of both values, we attempted varying permutations of the second set of indices to identify the optimal alignment. The resulting correlations, referred to as 'R-indices-correlation', represent the means of these two measurements and serve as a quantitative indicator of the agreements between r^1 and r^2.

Method 7: replication experiments

We first assessed the replicability of five dimensions independently within male and female groups. Specifically, we divided the original training samples into 4,176 male (528 REF and 3,648 TAR) and 5,966 female (622 REF and 5,344 TAR) participants. Subsequently, Surreal-GAN was separately retrained on each population with the same set of hyperparameters to derive sex-specific dimensions.

To further test the reproducibility of the identified five dimensions, we retrained the model using a completely independent training set and performed comparisons. Specifically, we resampled 1,000 pre-aging and 4,818 aging participants not included in the original training sets. A maximum of 500 individuals from each study were included in the REF group (pre-aging group), while up to 2,000 individuals were allocated to the TAR group (aging group). The same procedure introduced in Method 4 was used to derive R-indices for both the replication training set and all elderly participants.

Method 8: semi-synthetic experiments

We performed several semi-synthetic experiments to evaluate the robustness of the improved Surreal-GAN model across various scenarios and to validate our hyperparameter selection strategy (Method 4). For semi-synthetic data construction, we followed the semisynthetic dataset construction approach from the Surreal-GAN paper, partitioning 1,779 individuals who were CN (age < 70 years) into a REF group (579 individuals) and a Pseudo-TAR group (1,200 individuals). For each participant in the Pseudo-TAR group (denoted as the i_{th} participant), a three-dimensional pattern severity vector (s_i) was sampled from a multivariate uniform distribution ($\mathrm{s}_\mathrm{i} \sim \mathrm{U}[0,\ 1]$), representing the ground truth. We introduced three atrophy patterns to the 1,200 Pseudo-TAR individuals based on the sampled severity vectors, each involving distinct regions. Unlike the Surreal-GAN paper, we also simulated different levels of correlations among three dimensions. (1) For singular correlation, correlations of 0.2, 0.4 and 0.6 were only simulated between two dimensions, respectively. (2) For comprehensive batch size TAR sample size correlation; correlations of 0.2, 0.4 and 0.6 were simulated across all three dimensions.

First, the atrophy rate among the target population has been an important factor affecting the reliability of the original Surreal-GAN model as introduced in Yang et al.[10]. We re-evaluated the improved Surreal-GAN's performances across various rates of brain changes. We constructed semi-synthetic datasets with the same three simulated patterns but distinct ranges of simulated brain changes: 0–30%, 0–20% and 0–15%. We not only evaluated each model's performance on the datasets it was initially trained on but also assessed its efficacy on the other two datasets with different degrees of atrophy. R-indices-correlations were calculated between the derived R-indices and the ground truth, s, to quantify their respective performances (Supplementary Table 2). Additionally, to better understand atrophy rates within the context of brain aging, we used the mean and s.d. estimated from the reference group to compute the z-scores of regions of interest (ROIs) across all datasets, including the three semi-synthetic ones and the real data from the aging population (Supplementary Table 3).

Second, to assess the improved Surreal-GAN's ability to capture associations among underlying pathologies, we compared two models using semi-synthetic datasets with 0–20% simulated brain changes. Both the improved and original Surreal-GAN models were repetitively trained 50 times on each semi-synthetic dataset with different levels of correlation.

Finally, we tested validity of the hyperparameter selection strategy mentioned in Method 4, using the semi-synthetic dataset with a maximum of 15% simulated atrophy rate. Due to the difficulty of the problem, there are larger variations in representation accuracies across different hyperparameters, rendering this dataset particularly suitable for evaluating our hyperparameter selection strategy (Supplementary Table 1).

Method 9: image processing and harmonization

A fully automated pipeline was applied to process the T1-weighted MRIs. All MRIs were first corrected for intensity inhomogeneities.[86] A multi-atlas skull-stripping algorithm was applied to remove extracranial material.[87] Subsequently, 139 anatomical ROIs were identified in gray matter (119 ROIs) and white matter (20 ROIs) using a multi-atlas label fusion method[88]. These multi-atlas segmentation methods offer robustness against errors in individual deformations due to their reliance on multiple anatomical reference images (atlases) that are individually warped to the target image using deformable registration.

Due to the large sample size, we implemented a two-step semi-automated quality control process for the aforementioned processing, applied separately to different data batches from various studies. First, an automated ranking system assessed the distribution of segmented ROI volumes within each batch, transforming the values into z-scores and calculating a cumulative ranking score. This identified and ranked images that deviated most from the expected ranges. A predefined number of ranked images were reviewed manually using MRISnapshot, an in-house quality control tool accessible at https://github.com/CBICA/MRISnapshot. MRISnapshot's user-friendly web interface allows for efficient navigation and tagging of large datasets by displaying overlays of selected labels directly on specific image slices in a configurable way.

We further merged symmetric ROIs from the left and right hemispheres, resulting in 72 ROI volumes used as features for the Surreal-GAN model. Voxel-wise regional volumetric maps for gray matter and white matter tissues (referred to as RAVENS)[89], were computed by spatially aligning skull-stripped images to a single-subject brain template using a registration method[90]. WMH volumes were calculated through a deep-learning-based segmentation method[88] built upon U-Net architecture[91], using inhomogeneity-corrected and co-registered FLAIR and T1-weighted images. Site-specific mean and variance were estimated with an extensively validated statistical harmonization method[92] in the CN population and applied to the entire population while controlling for covariates.

Method 10: environmental/lifestyle factors and clinical variables

From the UKBB study, based on the variables analyzed by Tian et al.[5], we selected all 120 variables that also exist in our UKBB data collection. These variables indicate individual differences in early life experience (for example, birth weight and age at live births), lifestyle (for example, smoking and alcohol consumption), social recreation (for example, frequency of friend or family visits and time spent watching television), psychological condition (for example, mood status), local environmental exposures (for example, coastal proximity and air pollution) and general health. Several variables were recoded for more convenient interpretation, as introduced by Tian et al.[5]. The full list of selected variables with corresponding UKBB data coding is provided in Supplementary Data 4.

Method 11: cognitive, clinical, CSF and plasma biomarkers

We included CSF and plasma biomarkers provided by ADNI, as well as cognitive test scores provided by nine studies. For ADNI, all measures were downloaded from the LONI website (http://adni.loni.ucla.edu). Detailed methods for CSF measurements of β-amyloid (Aβ) and phospho-tau (p-tau) are described by Hansson et al.[93] Other CSF and plasma biomarkers were measured using the multiplex xMAP Luminex platform, with details described in 'Biomarkers Consortium ADNI Plasma Targeted Proteomics Project – Data Primer' (available at http://adni.loni.ucla.edu). The ADNI study has previously validated several composite cognitive scores across several domains, including ADNI-MEM[94], ADNI-EF[95], the visuospatial functioning composite (ADNI-VS)[96] and the language composite (ADNI-LAN)[96]. ADNI-MEM is based on components from the Rey Auditory Verbal Learning Test, Alzheimer's Disease Assessment Scale–Cognitive Subscale (ADAS-Cog) and the Mini-Mental State Exam (MMSE). ADNI-EF uses animal and vegetable category fluency, Trail-Making A and B, Digit Span Backwards, Digit Symbol Substitution from the Wechsler Adult Intelligence Scale-Revised and circle, symbol, numbers, hands and time items from a clock drawing task. ADNI-LAN uses animal and vegetable category fluency, the Boston Naming total, MMSE language elements, following commands/object naming/

ideational practice from ADAS-Cog and Montreal Cognitive Assessment language elements, including letter fluency, naming and repeating tasks. ADNI-VS is constructed based on clock copying with five different scored elements, the copying interlocking pentagons element of the MMSE and a constructional praxis item of ADAS-Cog. Beyond these composite scores that encompass varied tests, we also included four other cognitive scores broadly available among most (at least eight) studies (Supplementary Data 9). Education attainment variables were collected using different criteria in different studies. Thus, we mapped them to five ordinal levels, as detailed in Supplementary Data 10.

Method 12: health outcomes

Based on the self-report (field ID 20002) and healthcare records (field IDs 41270 and 41271) from the UKBB study, we defined the patient group of 14 different chronic diseases, including MCI/dementia, stroke, multiple sclerosis, hypertensive diseases, diabetes, depression, bipolar disorder, schizophrenia, Parkinson’s disease, COPD, osteoarthritis, CKD, osteoporosis and ischemic heart disease. Supplementary Data 11 lists noncancer illnesses, International Classification of Diseases 9 and 10 codes related to each of the 14 disease categories. Additionally, the MCI/dementia category also includes 2,550 participants from six other studies, including ADNI, AIBL, BIOCARD, BLSA, OASIS and PENN. The schizophrenia category includes an additional 71 participants from PHENOM (Psychosis Heterogeneity Evaluated via Dimensional Neuroimaging). Individuals diagnosed with more than one disease category were assigned to multiple groups. Also, 2,971 UKBB participants without any of the 14 diseases were categorized as a HC group. Mortality data released on 4 March 2021, from UKBB (field ID 40000) were used for analyses of the risk of mortality. The dates of death were determined through data linkages to national death registries in the UK (documentation at https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/DeathLinkage.pdf). Overall, 397 participants were confirmed dead after the acquisitions of brain MRIs at baseline.

Method 13: statistical analyses on baseline variables

We first calculated Pearson correlations between R-indices and chronological age, as well as among all R-indices. Partial correlation has been used to measure associations between two variables while controlling for the effects of a set of covariates. Given that age and sex may demonstrate significant correlations with both R-indices and other analyzed variables, we utilized partial correlation, with age and sex corrected as covariates, to examine the additional variations in all other baseline variables explained by R-indices across various age and sex groups. When comparing HC with each disease group in R-indices, we first adjusted for age and sex effects through multiple linear regression ($\mathrm{R} = \beta_{disease}\mathbf{x}_{disease} + \beta_{age}\mathbf{x}_{age} + \beta_{sex}\mathbf{x}_{sex} + b$) and then calculated the Cohen’s d effect size between the two groups. Mathematically, partial correlation is equivalent to multiple regression with controlled covariates included as right-side variables, so they share the same P values, with partial correlation being a standardized version of the β coefficient. For variables from multiple studies, the categorized study information was also incorporated as additional covariates. The false discovery rate was controlled at 5% using the Benjamini–Hochberg procedure for CSF/plasma biomarkers due to the small sample size. Bonferroni correction was used to control the family-wise error elsewhere.

Through the Python package Nilearn[97], voxel-based morphometry analyses were used for testing associations between voxel-wise tissue density and each R-index, adjusting for age, sex, ICV and the remaining four R-indices. For voxel-based morphometry, The false discovery rate was controlled at 0.1% using the Benjamini–Hochberg procedure.

Method 14: association with future risk of disease progression or mortality

Toevaluate the association between each R-index and future progression from CN to MCI or from MCI to dementia, we included 5,807 participants with longitudinal diagnoses available, among whom 4,777 were diagnosed as CN and 1,030 were diagnosed as MCI at baseline. We employed a Cox proportional hazard model while adjusting for covariates such as age and sex to test the associations. The covariates were included as additional right-side variables in the model, so the hazard function has the following form: $h(t \mid R,\ \mathbf{x}_{age},\ \mathbf{x}_{sex}) = h_0(t)\exp(\beta_R R + \beta_{age}\mathbf{x}_{age} + \beta_{sex}\mathbf{x}_{sex}\,.)$The HR, $\exp(\beta_R)$, was calculated and reported as the effect size measure that indicates the influence ofeach R-indexonthe riskofdisease progression. Further, to quantitatively assess prognostic performances with R-indices as features, we progressively added the most predictive R-indices to the Cox model to understand its optimal performance. The concordance index (C-index) was utilized to quantify the performance of risk prediction in a 100 repetition of 20% holdout cross validation. Using the same pipeline, we also evaluated the association between each R-index and the risk of mortality using all 38,675 participants from UKBB, with 397 confirmed dead after baseline assessments.

Method 15: genetic analyses

We used the imputed genotype data for all genetic analyses. Our quality check pipeline resulted in 32,829 participants with European ancestry and 8,469,833 SNPs. To summarize, we excluded individuals or genetic variants based on the following exclusion criteria: (1) related individuals (up to second-degree) identified through family relationship[98]; (2) duplicated variants; (3) individuals whose self-acknowledged sex did not match genetically identified sex; (4) individuals with more than 3% of missing genotypes; (5) variants with minor allele frequency of less than 1%; (6) variants with larger than 3% missing genotyping rate; and (7) variants that failed the Hardy–Weinberg test at 1×10^{-10}. To adjust for population stratification[99], we derived the first 40 genetic principal components using the FlashPCA software[100]. Details of the genetic quality check protocol are described elsewhere[35,26].

15a GWAS.—For GWAS, we ran a linear regression using Plink[101] for each R-index, controlling for confounders of age, sex, age–sex interaction, age-squared, age-squared–sex interaction, total ICV and the first 40 genetic principal components. We adopted the genome-wide *P* value threshold (5×10^{-8}) and annotated independent genetic signals considering linkage disequilibrium.

15b Phenome-wide association queries for the identified loci in GWAS catalog.—We queried the significant independent SNPs within each locus in the EMBL-EBI GWAS catalog (query date, 2 June 2023, via FUMA[102] v.1.5.4) to determine their previously identified associations with any other traits. We further grouped these traits into 11 categories for visualization and interpretation.

15c Gene-set enrichment analyses.—We conducted gene-set enrichment analyses using gene sets from the MsigDB database (v.6.2). Bonferroni correction was performed for all tested genes ($P < 2.64 \times 10^{-6}$) and gene sets ($P < 4.68 \times 10^{-6}$). All other parameters were set by default in FUMA.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All derived R-indices in this study are available at Supplementary Data 12, indexed by participant ID in the iSTAGING study. Additional raw imaging and clinical data used in this study were provided by several individual studies via data-sharing agreements, which do not include permission for us to further share the data. Investigators must apply to the source data providers to access additional data and match their subject IDs to those used in this study under the current protocol (primarily for UKBB). Data from ADNI are available from the ADNI database (adni.loni.usc.edu) upon registration and compliance with the data usage agreement. Data from the UKBB are available upon request from the UKBB website (https://www.ukbiobank.ac.uk/). Data from the BLSA study are available upon request at https://www.blsa.nih.gov/how-apply. Data from the AIBL study are available upon request at https://aibl.org.au/. Data from the OASIS study are available upon request at https://www.oasis-brains.org/. Data requests for BIOCARD, PENN, WRAP, CARDIA, SHIP and WHIMS datasets should be directed to M.S.A., D.A.W., S.C.J., L.J.L., K.W. and M.A.E., respectively. As soon as access to the source studies is obtained, investigators can match our derived R-indices to the rest of the data from these studies. Further assistance in matching the R-indices can be requested from the corresponding author, C.D., at Christos.Davatzikos@pennmedicine.upenn.edu, with responses typically provided within 2 weeks. Moreover, we are actively following protocols to upload our derived measures to the UKBB and ADNI websites, making them directly accessible to investigators who obtain access to those studies. The pretrained model for deriving R-indices in this study is available at https://github.com/zhijian-yang/SurrealGAN/blob/main/pretrained_models/brain_aging_5rindices/. Researchers can derive R-indices on their own datasets by following the data processing pipeline outlined in Method 9 and the model application process in Method 4, along with the example script on the same GitHub repository. The GWAS summary statistics are publicly available at https://labs-laboratory.com/medicine.

Code availability

The software Surreal-GAN is available as a published PyPI package. Detailed information about software installation, usage and license can be found at https://pypi.org/project/SurrealGAN/0.1.1/. Custom code can be found at https://github.com/zhijian-yang/

SurrealGAN. The deep-learning models are currently being integrated into NiChart (neuroimagingchart.com), enabling researchers to quickly derive R-indices for out-of-domain structural brain MRI scans.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Zhijian Yang[1,2,3], Junhao Wen[4], Guray Erus[1], Sindhuja T. Govindarajan[1], Randa Melhem[1], Elizabeth Mamourian[1], Yuhan Cui[1], Dhivya Srinivasan[1], Ahmed Abdulkadir[5], Paraskevi Parmpi[1], Katharina Wittfeld[6], Hans J. Grabe[6,7], Robin Bülow[8], Stefan Frenzel[6], Duygu Tosun[9], Murat Bilgel[10], Yang An[10], Dahyun Yi[11], Daniel S. Marcus[12], Pamela LaMontagne[12], Tammie L. S. Benzinger[12], Susan R. Heckbert[13], Thomas R. Austin[13], Shari R. Waldstein[14], Michele K. Evans[15], Alan B. Zonderman[15], Lenore J. Launer[16], Aristeidis Sotiras[17], Mark A. Espeland[18], Colin L. Masters[19], Paul Maruff[19], Jurgen Fripp[20], Arthur W. Toga[21], Sid O'Bryant[22], Mallar M. Chakravarty[23], Sylvia Villeneuve[24], Sterling C. Johnson[25], John C. Morris[26], Marilyn S. Albert[27], Kristine Yaffe[28], Henry Völzke[29], Luigi Ferrucci[30], R. Nick Bryan[31], Russell T. Shinohara[1,32], Yong Fan[1], Mohamad Habes[33], Paris Alexandros Lalousis[34], Nikolaos Koutsouleris[34,35], David A. Wolk[36], Susan M. Resnick[10], Haochang Shou[1,32], Ilya M. Nasrallah[1,31], Christos Davatzikos[1,✉]

Affiliations

[1] Artificial Intelligence in Biomedical Imaging Laboratory (AIBIL), Center for and Data Science for Integrated Diagnostics (AI2D), Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

[2] Graduate Group in Applied Mathematics and Computational Science, University of Pennsylvania, Philadelphia, PA, USA.

[3] GE Healthcare, Bellevue, WA, USA.

[4] Laboratory of AI and Biomedical Science (LABS), Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, USA.

[5] Laboratory for Research in Neuroimaging, Department of Clinical Neurosciences, Lausanne University Hospital (CHUV) and University of Lausanne, Lausanne, Switzerland.

[6] Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany.

[7] Site Rostock/Greifswald, German Center for Neurodegenerative Diseases (DZNE), Greifswald, Germany.

[8] Institute of Diagnostic Radiology and Neuroradiology, University of Greifswald, Greifswald, Germany.

[9]Department of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco, CA, USA.

[10]Laboratory of Behavioral Neuroscience, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA.

[11]Institute of Human Behavioral Medicine, Medical Research Center Seoul National University, Seoul, Republic of Korea.

[12]Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, MO, USA.

[13]Cardiovascular Health Research Unit and Department of Epidemiology, University of Washington, Seattle, WA, USA.

[14]Department of Psychology, University of Maryland, Baltimore County, Baltimore, MD, USA.

[15]Health Disparities Research Section, Laboratory of Epidemiology and Population Sciences, NIA/NIH/IRP, Baltimore, MD, USA.

[16]Neuroepidemiology Section, Intramural Research Program, National Institute on Aging, Bethesda, MD, USA.

[17]Department of Radiology and Institute for Informatics, Data Science & Biostatistics, Washington University in St. Louis, St. Louis, MO, USA.

[18]Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, NC, USA.

[19]Florey Institute, The University of Melbourne, Parkville, Victoria, Australia.

[20]CSIRO Health and Biosecurity, Australian e-Health Research Centre CSIRO, Brisbane, Queensland, Australia.

[21]Laboratory of Neuro Imaging, USC Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, USA.

[22]Institute for Translational Research University of North Texas Health Science Center, Fort Worth, TX, USA.

[23]Computational Brain Anatomy (CoBrA) Laboratory, Cerebral Imaging Center, Douglas Mental Health University Institute, McGill University, Verdun, Quebec, Canada.

[24]McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada.

[25]Wisconsin Alzheimer’s Institute, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA.

[26]Knight Alzheimer Disease Research Center, Dept of Neurology, Washington University School of Medicine, St. Louis, MO, USA.

[27]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

[28]Departments of Neurology, Psychiatry and Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA.

[29]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany.

[30]Translational Gerontology Branch, Longitudinal Studies Section, National Institute on Aging, National Institutes of Health, MedStar Harbor Hospital, Baltimore, MD, USA.

[31]Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA.

[32]Penn Statistics in Imaging and Visualization Center, Department of Biostatistics, Epidemiology, & Informatics, University of Pennsylvania, Philadelphia, PA, USA.

[33]Neuroimage Analytics Laboratory and Biggs Institute Neuroimaging Core, Glenn Biggs Institute for Neurodegenerative Disorders, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA.

[34]Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King’s College London, London, UK.

[35]Section for Precision Psychiatry, Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University Munich, Munich, Germany.

[36]Department of Neurology, University of Pennsylvania, Philadelphia, PA, USA.

Acknowledgements

Data used in this study are part of the iSTAGING study (representative, C. Davatzikos), the Preclinical AD Consortium (M. S. Albert), ADNI (M. W. Weiner) and the BLSA (S. M. Resnick). The iSTAGING consortium is a multi-institutional effort funded by the National Institute on Aging (NIA) by RF1 AG054409. The BLSA neuroimaging study is funded by the Intramural Research Program, NIA, National Institutes of Health (NIH) and by HHSN271201600059C. The BIOCARD study is in part supported by NIH grant U19-AG033655. SHIP is part of the Community Medicine Research net of the University of Greifswald, which is funded by the Federal Ministry of Education and Research (grant nos. 01ZZ9603, 01ZZ0103 and 01ZZ0403), the Ministry of Cultural Affairs and the Social Ministry of the Federal State of Mecklenburg-West Pomerania. MRI scans in SHIP-START and SHIP-TREND have been supported by a joint grant from Siemens Healthineers and the Federal State of Mecklenburg-West Pomerania. The Women’s Health Initiative was funded by the National Heart, Lung and Blood Institute of the NIH, US Department of Health and Human Services. Contracts HHSN268200464221C and N01-WH-4–4221 provided additional support. The WHIMS (M.A.E.) was funded in part by Wyeth Pharmaceuticals. The HANDLS study is supported by the NIA Intramural Research Program, NIH, Project ZIA-AG000513. The HANDLS Scan substudy is supported by NIH grants 1RO1AG034161, 2P30AG028747–14S1 and 1R56AG064088–01A1. The HABS-HD project is funded by grants from the NIA: R01AG054073 and R01AG058533. HABS-HD multiple principal investigators were S. E. O’Bryant, K. Yaffe, A. Toga, R. Rissman and L. Johnson; and the HABS-HD investigators were M. Braskie, K. King, J. R Hall, M. Petersen, R. Palmer, R. Barber, Y. Shi, F. Zhang, R. Nandy, R. McColl, D. Mason, B. Christian, N. Philips, S. Large, J. Lee, B. Vardarajan, M. Rivera Mindt, A. Cheema, L. Barnes, M. Mapstone, A. Cohen, A. Kind, O. Okonkwo, R. Vintimilla, Z. Zhou, M. Donohue, R. Raman, M. Borzage, M. Mielke, B. Ances, G. Babulal, J. Llibre-Guerra, C. Hill and R. Vig. Other supporting grants include 5U01AG068057–02 and 1U24AG074855–01. S.R.H. and J.C.M. are supported by multiple grants and contracts from NIH. A.A. was supported by grants 191026 and 206795 from the Swiss National Science Foundation. R.T.S. was supported by grants R01MH123550 and R01MH112847. J.C.M. was supported by NIH ACS grant, P01AG026276. Y.F. was supported by grant R01AG066650. M.H. was supported by the National Institute of Health (NIH) grant number 1R01AG080821. This research has been conducted using the UKBB Resource under application no. 35148. Data used in the preparation of this article were in part obtained from the ADNI database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. ADNI is funded by the NIA, the National Institute of Biomedical Imaging and Bioengineering and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica; Biogen; Bristol-Myers Squibb; CereSpir; Cogstate; Eisai; Elan Pharmaceuticals; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche and its affiliated company Genentech; Fujirebio; GE Healthcare; IXICO; Janssen Alzheimer Immunotherapy Research & Development; Johnson & Johnson Pharmaceutical Research & Development; Lumosity; Lundbeck; Merck & Co; Meso Scale Diagnostics; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. Z.Y. had full access to all the data in the study and is responsible for the integrity of the data and the accuracy of the data analysis.

Competing interests

H.J.G. has received travel grants and speaker's honoraria from Fresenius Medical Care, Neuraxpharm, Servier and Janssen Cilag as well as research funding from Fresenius Medical Care. R.T.S. received consulting income from Octave Bioscience and has received compensation for scientific reviewing from the American Medical Association. T.L.S.B. has received investigator-initiated research funding from the NIH, the Alzheimer's Association, the Foundation at Barnes-Jewish Hospital, Siemens Healthineers and Avid Radiopharmaceuticals (a wholly owned subsidiary of Eli Lilly and Company). She participates as a site investigator in clinical trials sponsored by Eli Lilly and Company, Biogen, Eisai, Jaansen and Roche. She has served as a paid and unpaid consultant to Eisai, Siemens, Biogen, Janssen and Bristol-Myers Squibb. The other authors declare no competing interests.

References

1. Peters R Ageing and the brain. Postgrad. Med J 82, 84–88 (2006).
2. Davatzikos C, Xu F, An Y, Fan Y & Resnick SM Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the SPARE-AD index. Brain 132, 2026–2035 (2009).
3. Dubois B et al. Preclinical Alzheimer's disease: definition, natural history, and diagnostic criteria. Alzheimers Dement 12, 292–323 (2016).
4. Davatzikos C Machine learning in neuroimaging: progress and challenges. Neuroimage 197, 652–656 (2019).
5. Tian YE et al. Heterogeneous aging across multiple organ systems and prediction of chronic disease and mortality. Nat. Med 29, 1221–1231 (2023).
6. Habes M et al. Advanced brain aging: relationship with epidemiologic and genetic risk factors, and overlap with Alzheimer disease atrophy patterns. Transl. Psychiatry 6, e775 (2016).
7. Yang Z et al. A deep learning framework identifies dimensional representations of Alzheimer's disease from brain structure. Nat. Commun 12, 7065 (2021).
8. Zhang X et al. Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. Proc. Natl Acad. Sci. USA 113, E6535–e6544 (2016).
9. Wen J et al. Multi-scale semi-supervised clustering of brain images: deriving disease subtypes. Med. Image Anal 75, 102304 (2022).
10. Yang Z, Wen J & Davatzikos C Surreal-GAN: Semi-Supervised Representation Learning via GAN for uncovering heterogeneous disease-related imaging patterns. International Conference on Learning Representations (ICLR, 2022).
11. Habes M et al. The brain chart of aging: Machine-learning analytics reveals links between brain aging, white matter disease, amyloid burden, and cognition in the iSTAGING consortium of 10,216 harmonized MR scans. Alzheimers Dement 17, 89–102 (2021).
12. Cox SR et al. Ageing and brain white matter structure in 3,513 UK Biobank participants. Nat. Commun 7, 13629 (2016).
13. Hedden T & Gabrieli JDE Insights into the ageing mind: a view from cognitive neuroscience. Nat. Rev. Neurosci 5, 87–96 (2004).

14. Wang MC, Shah NS, Carnethon MR, O'Brien MJ & Khan SS Age at diagnosis of diabetes by race and ethnicity in the United States from 2011 to 2018. JAMA Intern. Med 181, 1537–1539 (2021).
15. Huang X, Lee K, Wang MC, Shah NS & Khan SS Age at diagnosis of hypertension by race and ethnicity in the US from 2011 to 2020. JAMA Cardiol 7, 986–987 (2022).
16. Abbott A Dementia: a problem for our age. Nature 475, S2–S4 (2011).
17. Dwyer DB et al. Psychosis brain subtypes validated in first-episode cohorts and related to illness remission: results from the PHENOM consortium. Mol. Psychiatry 28, 2008–2017 (2023).
18. Stark K & Massberg S Interplay between inflammation and thrombosis in cardiovascular pathology. Nat. Rev. Cardiol 18, 666–682 (2021).
19. Rose-John S, Winthrop K & Calabrese L The role of IL-6 in host defence against infections: immunobiology and clinical implications. Nat. Rev. Rheumatol 13, 399–409 (2017).
20. Dutta G, Barber DS, Zhang P, Doperalski NJ & Liu B Involvement of dopaminergic neuronal cystatin C in neuronal injury-induced microglial activation and neurotoxicity. J. Neurochem 122, 752–763 (2012).
21. Buniello A et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 47, D1005–d1012 (2019).
22. Zhao B et al. Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. Nat. Genet 51, 1637–1644 (2019).
23. Zhao B et al. Large-scale GWAS reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive and mental health traits (n = 17,706). Mol. Psychiatry 26, 3943–3955 (2021).
24. Seshadri S et al. Genetic correlates of brain aging on MRI and cognitive test measures: a genome-wide association and linkage analysis in the Framingham study. BMC Med. Genet 8, S15 (2007).
25. Leonardsen EH et al. Genetic architecture of brain age and its causal relations with brain and mental disorders. Mol. Psychiatry 28, 3111–3120 (2023).
26. Wen J et al. The genetic architecture of multimodal human brain age. Nat. Commun 15, 2604 (2024).
27. Chauhan G et al. Association of Alzheimer's disease GWAS loci with MRI markers of brain aging. Neurobiol. Aging 36, 1765.e1767–1765.e1716 (2015).
28. Binnewies J et al. Lifestyle-related risk factors and their cumulative associations with hippocampal and total grey matter volume across the adult lifespan: a pooled analysis in the European Lifebrain consortium. Brain Res. Bull 200, 110692 (2023).
29. Fotuhi M, Do D & Jack C Modifiable factors that alter the size of the hippocampus with ageing. Nat. Rev. Neurol 8, 189–202 (2012).
30. Kapasi A, DeCarli C & Schneider JA Impact of multiple pathologies on the threshold for clinically overt dementia. Acta Neuropathol 134, 171–186 (2017).
31. Savva GM et al. Age, neuropathology, and dementia. N. Engl. J. Med 360, 2302–2309 (2009).
32. Dong A et al. Heterogeneity of neuroanatomical patterns in prodromal Alzheimer's disease: links to cognition, progression and biomarkers. Brain 140, 735–747 (2017).
33. Young AL et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with subtype and stage inference. Nat. Commun 9, 4273 (2018).
34. Chand GB et al. Two distinct neuroanatomical subtypes of schizophrenia revealed using machine learning. Brain 143, 1027–1038 (2020).
35. Wen J et al. Characterizing heterogeneity in neuroimaging, cognition, clinical symptoms, and genetics among patients with late-life depression. JAMA Psychiatry 79, 464–474 (2022).
36. Eavani H et al. Heterogeneity of structural and functional imaging patterns of advanced brain aging revealed via machine learning methods. Neurobiol. Aging 71, 41–50 (2018).
37. Habes M et al. White matter hyperintensities and imaging patterns of brain ageing in the general population. Brain 139, 1164–1179 (2016).
38. Skampardoni I et al. Genetic and clinical correlates of AI-based brain aging patterns in cognitively unimpaired individuals. JAMA Psychiatry 10.1001/jamapsychiatry.2023.5599 (2024).

39. Moonen JEF et al. Race, sex, and mid-life changes in brain health: Cardia MRI substudy. Alzheimers Dement 18, 2428–2437 (2022).

40. Nasrallah IM et al. Association of intensive vs standard blood pressure control with magnetic resonance imaging biomarkers of Alzheimer disease: secondary analysis of the SPRINT MIND randomized trial. JAMA Neurol 78, 568–577 (2021).

41. Schneider JA, Arvanitakis Z, Bang W & Bennett DA Mixed brain pathologies account for most dementia cases in community-dwelling older persons. Neurology 69, 2197–2204 (2007).

42. Pagani E et al. Regional brain atrophy evolves differently in patients with multiple sclerosis according to clinical phenotype. Am. J. Neuroradiol 26, 341–346 (2005).

43. Cagol A et al. Association of brain atrophy with disease progression independent of relapse activity in patients with relapsing multiple sclerosis. JAMA Neurol 79, 682–692 (2022).

44. Schoonheim MM, Broeders TAA & Geurts JJG The network collapse in multiple sclerosis: An overview of novel concepts to address disease dynamics. Neuroimage Clin 35, 103108 (2022).

45. Lee CU et al. Fusiform gyrus volume reduction in first-episode schizophrenia: a magnetic resonance imaging study. Arch. Gen. Psychiatry 59, 775–781 (2002).

46. Onitsuka T et al. Middle and inferior temporal gyrus gray matter volume abnormalities in chronic schizophrenia: an MRI study. Am. J. Psychiatry 161, 1603–1611 (2004).

47. Tremblay C et al. Brain atrophy progression in Parkinson's disease is shaped by connectivity and local vulnerability. Brain Commun 3, fcab269 (2021).

48. Kaur A et al. Structural and functional alterations of the temporal lobe in schizophrenia: a literature review. Cureus 12, e11177 (2020).

49. Gogolla N The insular cortex. Curr. Biol 27, R580–r586 (2017).

50. Oppenheimer SM, Gelb A, Girvin JP & Hachinski VC Cardiovascular effects of human insular cortex stimulation. Neurology 42, 1727–1732 (1992).

51. Fink JN et al. Insular cortex infarction in acute middle cerebral artery territory stroke: predictor of stroke severity and vascular lesion. Arch. Neurol 62, 1081–1085 (2005).

52. Craig AD How do you feel — now? The anterior insula and human awareness. Nat. Rev. Neurosci 10, 59–70 (2009).

53. Paulus MP & Stein MB An insular view of anxiety. Biol. Psychiatry 60, 383–387 (2006).

54. Brosch K et al. Reduced hippocampal gray matter volume is a common feature of patients with major depression, bipolar disorder, and schizophrenia spectrum disorders. Mol. Psychiatry 27, 4234–4243 (2022).

55. Alexandros Lalousis P et al. Transdiagnostic structural neuroimaging features in depression and psychosis: a systematic review. Neuroimage Clin 38, 103388 (2023).

56. Ribe AR et al. Long-term risk of dementia in persons with schizophrenia: a danish population-based cohort study. JAMA Psychiatry 72, 1095–1101 (2015).

57. de Lau LML, Schipper CMA, Hofman A, Koudstaal PJ & Breteler MMB Prognosis of Parkinson disease: risk of dementia and mortality: the rotterdam study. Arch. Neurol 62, 1265–1269 (2005).

58. Alzheimer's Association. 2016 Alzheimer's disease facts and figures. Alzheimers Dement 10.1016/j.jalz.2016.03.001 (2016).

59. Ten Kate M et al. Atrophy subtypes in prodromal Alzheimer's disease are associated with cognitive decline. Brain 141, 3443–3456 (2018).

60. Jack CR Jr. et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimers Dement 14, 535–562 (2018).

61. Schneider JA & Bennett DA Where vascular meets neurodegenerative disease. Stroke 41, S144–S146 (2010).

62. Smith AD Imaging the progression of Alzheimer pathology through the brain. Proc. Natl Acad. Sci. USA 99, 4135–4137 (2002).

63. Daviet R et al. Associations between alcohol consumption and gray and white matter volumes in the UK Biobank. Nat. Commun 13, 1175 (2022).

64. Topiwala A, Ebmeier KP, Maullin-Sapey T & Nichols TE Alcohol consumption and MRI markers of brain structure and function: Cohort study of 25,378 UK Biobank participants. Neuroimage Clin 35, 103066 (2022).

65. Karama S et al. Cigarette smoking and thinning of the brain's cortex. Mol. Psychiatry 20, 778–785 (2015).

66. Elbejjani M et al. Cigarette smoking and gray matter brain volumes in middle age adults: the CARDIA Brain MRI sub-study. Transl. Psych 10.1038/s41398-019-0401-1 (2019).

67. Kang J et al. Increased brain volume from higher cereal and lower coffee intake: shared genetic determinants and impacts on cognition and metabolism. Cereb. Cortex 32, 5163–5174 (2022).

68. de Lange AG et al. Population-based neuroimaging reveals traces of childbirth in the maternal brain. Proc. Natl Acad. Sci. USA 116, 22341–22346 (2019).

69. Aanes S, Bjuland KJ, Skranes J & Løhaugen GC Memory function and hippocampal volumes in preterm born very-low-birth-weight (VLBW) young adults. Neuroimage 105, 76–83 (2015).

70. Rodrigues D et al. Chronic stress causes striatal disinhibition mediated by SOM-interneurons in male mice. Nat. Commun 13, 7355 (2022).

71. Admon R et al. Striatal hypersensitivity during stress in remitted individuals with recurrent depression. Biol. Psychiatry 78, 67–76 (2015).

72. Lou C et al. Leveraging machine learning predictive biomarkers to augment the statistical power of clinical trials with baseline magnetic resonance imaging. Brain Commun 3, fcab264 (2021).

73. Goodfellow I et al. Generative adversarial networks. Adv. Neural Inf. Process. Syst 10.1145/3422622 (2014).

74. Chen X et al. InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets (NIPS, 2016).

75. Petersen RC et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. Neurology 74, 201–209 (2010).

76. Miller KL et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nat. Neurosci 19, 1523–1536 (2016).

77. Resnick SM et al. One-year age changes in MRI brain volumes in older adults. Cereb. Cortex 10, 464–472 (2000).

78. Resnick SM, Pham DL, Kraut MA, Zonderman AB & Davatzikos C Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain. J. Neurosci 23, 295–301 (2003).

79. Ellis KA et al. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. Int. Psychogeriatr 21, 672–687 (2009).

80. Soldan A et al. Relationship of medial temporal lobe atrophy, APOE genotype, and cognitive reserve in preclinical Alzheimer's disease. Hum. Brain Mapp 36, 2826–2841 (2015).

81. LaMontagne PJ et al. OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. Preprint at medRxiv 10.1101/2019.12.13.19014902 (2019).

82. Johnson SC et al. The Wisconsin registry for Alzheimer's prevention: a review of findings and current directions. Alzheimers Dement 10, 130–142 (2018).

83. Friedman GD et al. CARDIA: study design, recruitment, and some characteristics of the examined subjects. J. Clin. Epidemiol 41, 1105–1116 (1988).

84. Völzke H et al. Cohort profile: the study of health in Pomerania. Int. J. Epidemiol 40, 294–307 (2011).

85. Coker LH et al. Postmenopausal hormone therapy and subclinical cerebrovascular disease: the WHIMS-MRI Study. Neurology 72, 125–134 (2009).

86. Sled JG, Zijdenbos AP & Evans AC A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17, 87–97 (1998).

87. Doshi J, Erus G, Ou Y, Gaonkar B & Davatzikos C Multi-atlas skull-stripping. Acad. Radiol 20, 1566–1576 (2013).

88. Doshi J et al. MUSE: multi-atlas region segmentation utilizing ensembles of registration algorithms and parameters, and locally optimal atlas selection. NeuroImage 127, 186–195 (2016).

89. Davatzikos C, Genc A, Xu D & Resnick SM Voxel-based morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy. Neuroimage 14, 1361–1369 (2001).

90. Ou Y, Sotiras A, Paragios N & Davatzikos C DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting. Med. Image Anal 15, 622–639 (2011).

91. Ronneberger O, Fischer P & Brox T U-Net: convolutional networks for biomedical image segmentation. in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. (eds Navab N et al) 234–241 (Springer, 2015).

92. Pomponio R et al. Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. Neuroimage 208, 116450 (2020).

93. Hansson O et al. CSF biomarkers of Alzheimer's disease concord with amyloid-β PET and predict clinical progression: a study of fully automated immunoassays in BioFINDER and ADNI cohorts. Alzheimers Dement 14, 1470–1481 (2018).

94. Crane PK et al. Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). Brain Imaging Behav 6, 502–516 (2012).

95. Gibbons LE et al. A composite score for executive functioning, validated in Alzheimer's Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. Brain Imaging Behav 6, 517–527 (2012).

96. Choi SE et al. Development and validation of language and visuospatial composite scores in ADNI. Alzheimers Dement 6, e12072 (2020).

97. Abraham A et al. Machine learning for neuroimaging with scikit-learn. Front. Neuroinform 8, 14 (2014).

98. Manichaikul A et al. Robust relationship inference in genome-wide association studies. Bioinformatics 26, 2867–2873 (2010).

99. Price AL, Zaitlen NA, Reich D & Patterson N New approaches to population stratification in genome-wide association studies. Nat. Rev. Genet 11, 459–463 (2010).

100. Abraham G, Qiu Y & Inouye M FlashPCA2: principal component analysis of Biobank-scale genotype datasets. Bioinformatics 33, 2776–2778 (2017).

101. Purcell S et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet 81, 559–575 (2007).

102. Watanabe K, Taskesen E, van Bochoven A & Posthuma D Functional mapping and annotation of genetic associations with FUMA. Nat. Commun 8, 1826 (2017).

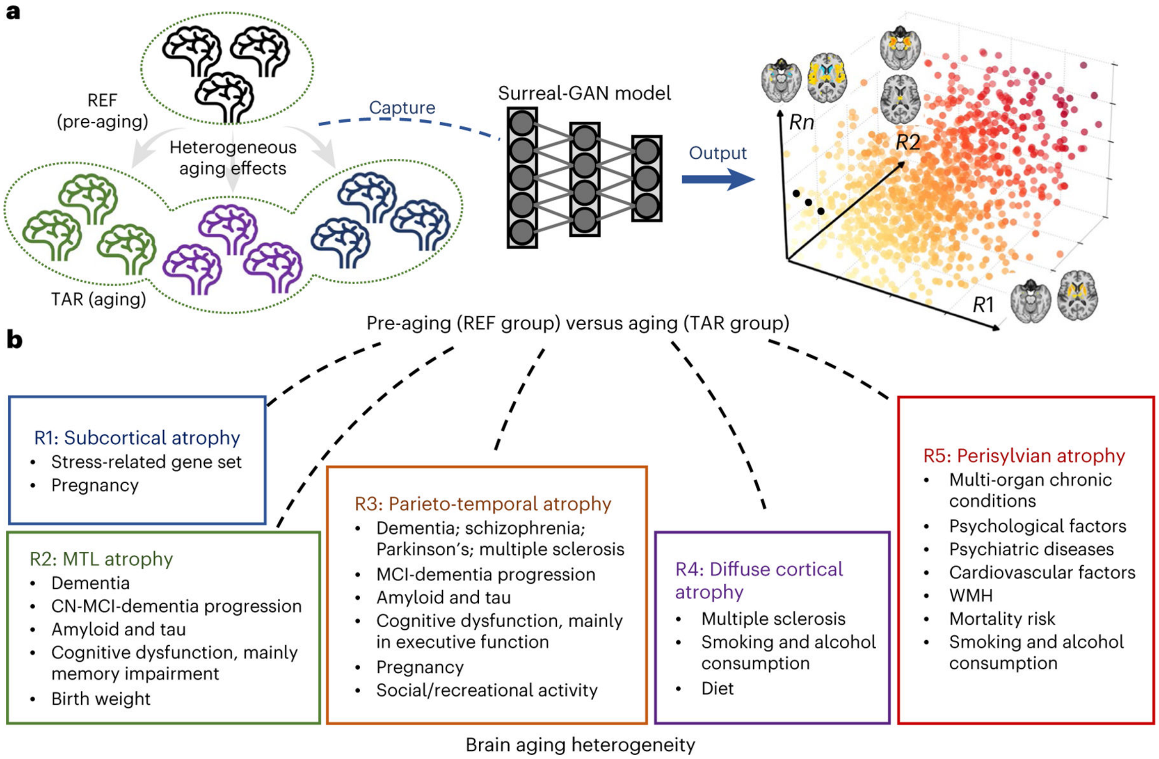


Fig. 1 |. Surreal-GAN disentangles brain aging heterogeneity through a dimensional representation approach.

a, The heterogeneous aging effects contribute to distinct alterations in human brain structures, leading to various brain change patterns. Surreal-GAN, a weakly supervised deep-learning approach utilizing generative learning, identifies patterns of brain change attributed to the aging process by capturing transformations from a REF population to a TAR population. It specifically represents the diversity of such brain change patterns in a given individual using multi-dimensional R-indices. These R-indices serve as indicators of and quantify the type and severity of distinct brain change patterns, which are presumed to reflect underlying neuropathological processes and their stages. **b**, In this study, to disentangle the neuroanatomical heterogeneity related to brain aging, we set the REF and TAR groups to pre-aging individuals (<50 years old) and all older adults (>50 years old), respectively. Surreal-GAN identifies five reproducible dimensions, each associated with distinct brain change patterns. Further statistical analyses uncover a range of influential factors associated with each dimension, encompassing pathological influences, lifestyle factors, life events and genetic variants.

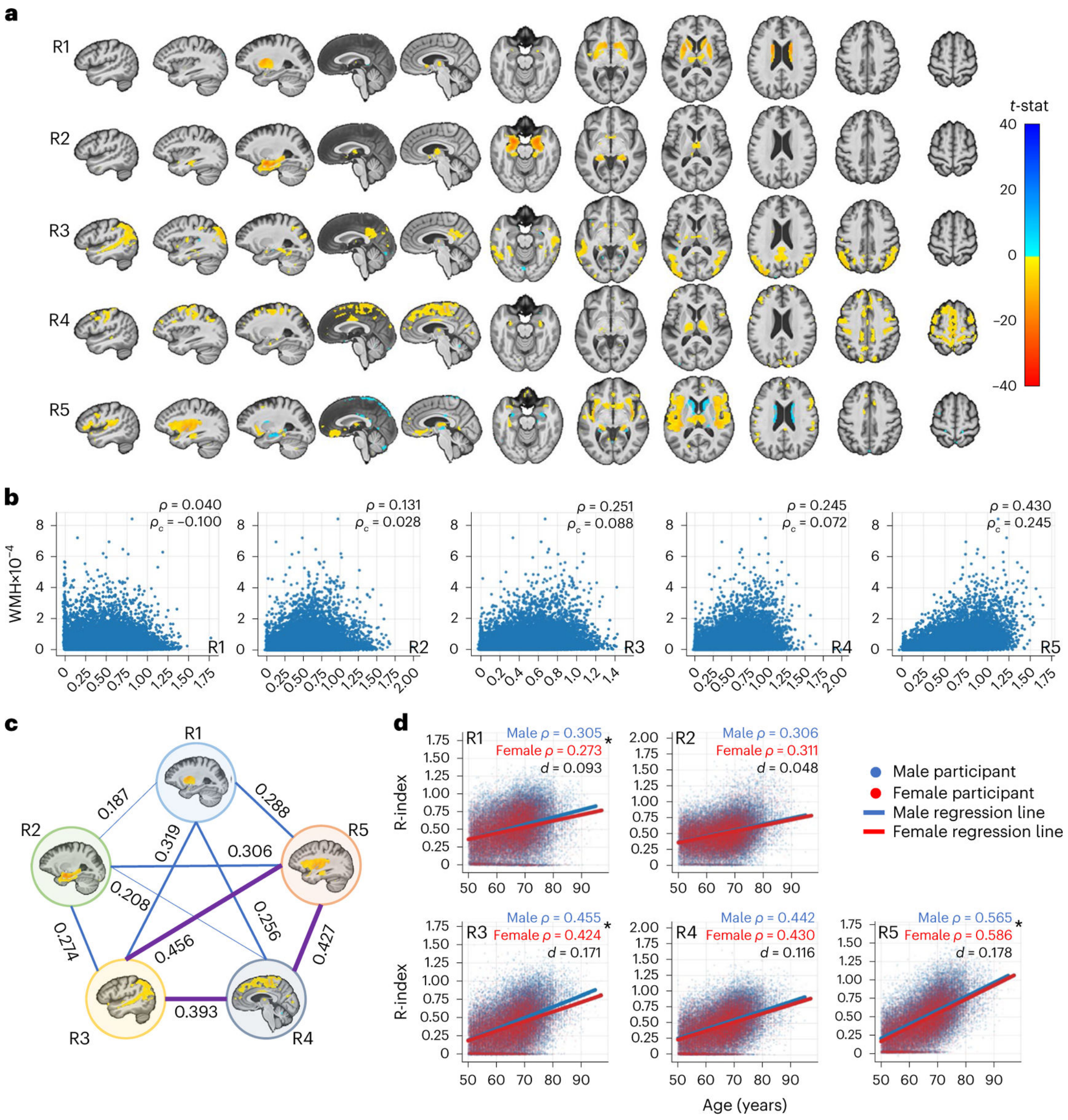


Fig. 2 |. Surreal-GAN identifies five dimensions of brain aging.

a, The severity of brain aging along five dimensions in each participant was quantified by the five R-indices (R1–R5), which revealed distinct patterns of associated gray matter atrophy. Characteristic patterns for each R-index are shown via voxel-wise *t*-tests performed for each R-index while adjusting for age, sex, intracranial volume (ICV) and the remaining four R-indices. False discovery rate (FDR) correction was performed to adjust multiple comparisons with a *P* value threshold of 0.001. **b**, The five R-indices show different levels of associations with WMH volumes. ρ_c and ρ denotes associations with and without adjusting for age and sex, respectively. R5 shows the strongest positive associations. **c**, The five R-indices demonstrate positive Pearson correlations with each other, with the strongest associations observed among R3, R4 and R5. **d**, The five R-indices exhibit significant positive associations with chronological age. Additionally, significant differences, marked by asterisks, were found between males and females in the correlations (ρ) between age and R1, R3 and R5. Moreover, adjusting for age, male and female groups show significant differences in distributions of R-indices, as shown by Cohen's d (Male > Female) values as effect sizes.

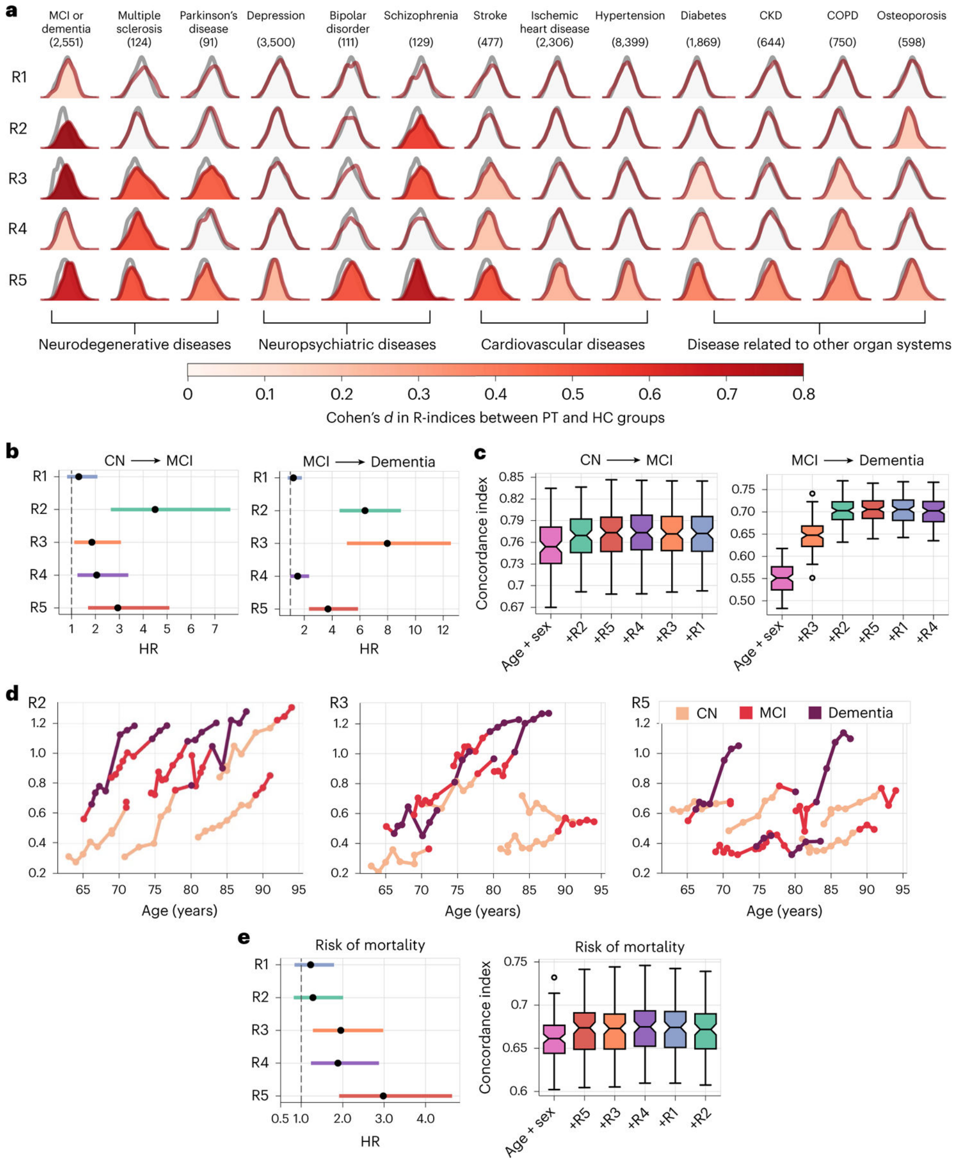
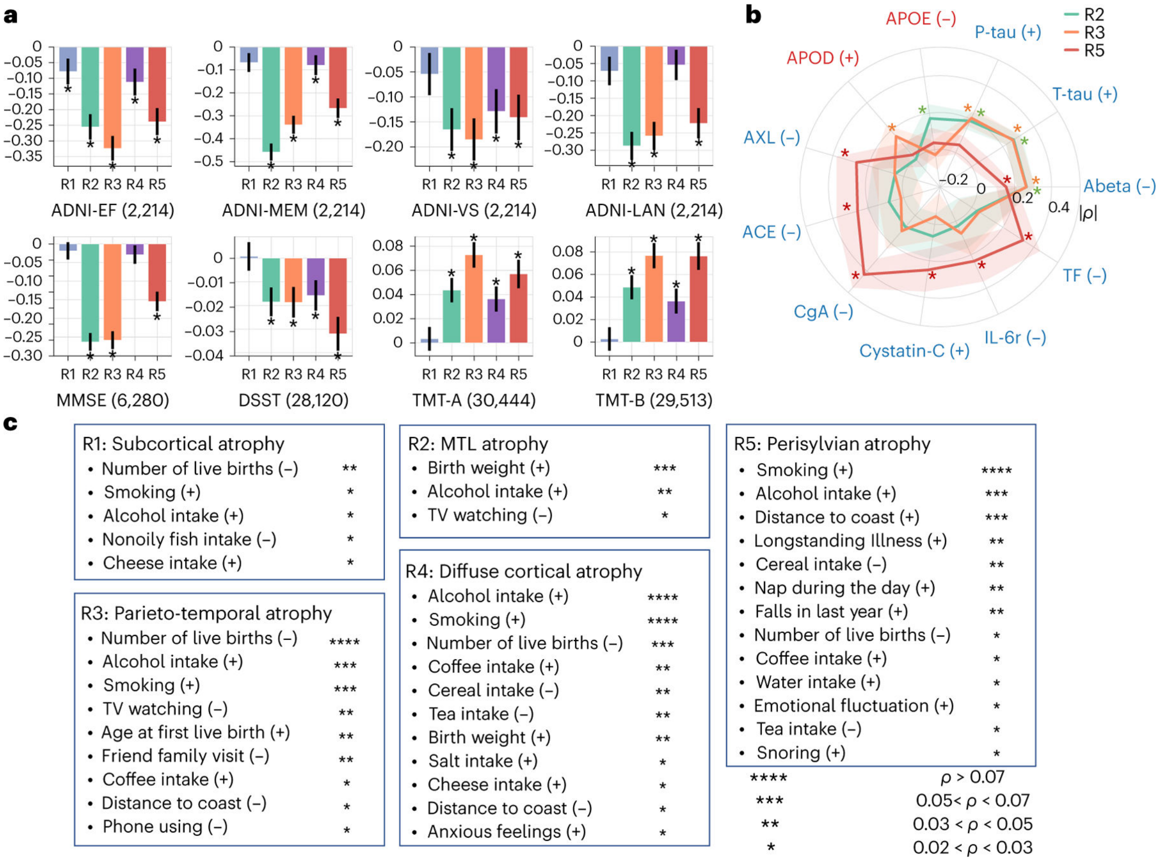


Fig. 3 |. R-indices are associated with chronic diseases, and MCI/Dementia progression, and the risk of mortality.

a, The distributions of R-indices are significantly different between the HC group and each patient (PT) group corresponding to one of the 13 chronic diseases, after adjusting for age and sex ($P < 6.6 \times 10^{-4}$, Bonferroni-corrected). Warmer colors denote larger Cohen's d (PT > HC). Distributions without color fill indicate no significant difference from the HC group (gray). **b**, Values of the R2–R5 indices exhibit associations with the risk of progression to MCI or dementia, as indicated by the corresponding HRs. Cox proportional hazard models were used for testing associations, adjusting for age and sex. **c**, The R-indices contribute to enhanced performance in predicting disease progression. Based on the significance of R-indices demonstrated in **b**, we progressively incorporated R-indices as features one by one when fitting the Cox proportional hazard model on participants over 60 years old at baseline. For each combination of features, 100 iterations of 20% holdout cross-validation were performed to derive concordance indices. **d**, The progression paths in R2, R3 and

R5 of eight representative participants transitioning either from CN to MCI or from MCI to dementia. Different colors represent the distinct diagnoses. **e**, The baseline R5 shows significant associations with the risk of mortality, with age and sex adjusted as covariates in Cox regression. Similar to **c**, R-indices were progressively included as features in cross-validation for mortality risk prediction among participants over 60 years old (error bar, 95% CI; center line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range; points, outliers).



R1: Subcortical atrophy

- Number of live births (–) **
- Smoking (+) *
- Alcohol intake (+) *
- Nonoily fish intake (–) *
- Cheese intake (+) *

R2: MTL atrophy

- Birth weight (+) ***
- Alcohol intake (+) **
- TV watching (–) *

R3: Parieto-temporal atrophy

- Number of live births (–) ****
- Alcohol intake (+) ***
- Smoking (+) ***
- TV watching (–) **
- Age at first live birth (+) **
- Friend family visit (–) **
- Coffee intake (+) *
- Distance to coast (–) *
- Phone using (–) *

R4: Diffuse cortical atrophy

- Alcohol intake (+) ****
- Smoking (+) ****
- Number of live births (–) ***
- Coffee intake (+) **
- Cereal intake (–) **
- Tea intake (–) **
- Birth weight (+) **
- Salt intake (+) *
- Cheese intake (+) *
- Distance to coast (–) *
- Anxious feelings (+) *

R5: Perisylvian atrophy

- Smoking (+) ****
- Alcohol intake (+) ***
- Distance to coast (+) ***
- Longstanding Illness (+) **
- Cereal intake (–) **
- Nap during the day (+) **
- Falls in last year (+) **
- Number of live births (–) *
- Coffee intake (+) *
- Water intake (+) *
- Emotional fluctuation (+) *
- Tea intake (–) *
- Snoring (+) *

****	$\rho > 0.07$
***	$0.05 < \rho < 0.07$
**	$0.03 < \rho < 0.05$
*	$0.02 < \rho < 0.03$

Fig. 4 |. Associations between R-indices and lifestyle, cognition and CSF/plasma biomarkers.

a, The five R-indices have distinct levels of association with different cognitive variables. Partial correlation (two-sided) was used for testing the associations between R-indices and cognitive scores, adjusting for age and sex. Additional site adjustments were performed for MMSE, DSST, TMT-A and TMT-B to account for the utilization of multi-site data. Significantly associated R-indices are marked by *($P < 1.25 \times 10^{-3}$, Bonferroni-corrected). Partial correlation coefficients are shown as centers of bar plots, with error bars representing 95% CIs. The sample sizes used to derive these coefficients are indicated next to the names of the cognitive variables. ADNI-MEM, ADNI-EF, ADNI-VS and ADNI-LAN are four ADNI composite cognitive scores related to memory, executive function, visuospatial functioning and language. **b**, Among the R-indices, R2, R3 and R5 have significant associations (marked by *) with 11 CSF/plasma biomarkers obtained from the ADNI study ($P < 6 \times 10^{-4}$). The CSF biomarkers are labeled in blue, and the plasma biomarkers are labeled in red. The radial graph presents the values (center) and 95% CI (error bands) of the correlation coefficients. For easier visualization, we invert the signs of negative coefficient (denoted by $|\rho|$) when making the plot. The '+' and '−' signs alongside the biomarker names indicate positive and negative correlations. Due to the small sample sizes, the Benjamini–Hochberg procedure was used for FDR correction **c**, The five R-indices show significant associations with a group of environmental/lifestyle factors and life events from the UKBB study ($P < 8.7 \times 10^{-5}$, Bonferroni-corrected). Partial correlation adjusting for age and sex was used, as in **a**. The number of '*' indicates correlation coefficients (legend). Positive and negative associations are denoted by ‘+’ and ‘−’ signs, respectively, adjacent to the factor names. For all three figures, two-sided *t*-tests were performed to test the significance of correlation coefficients. MMSE, Mini-Mental State Exam; DSST, Digit Symbol Substitution Test; TMT-A/B, Trail Making Test Part A/B; ADNI-VS, ADNI visuospatial functioning composite; ADNI-LAN, ADNI language composite.

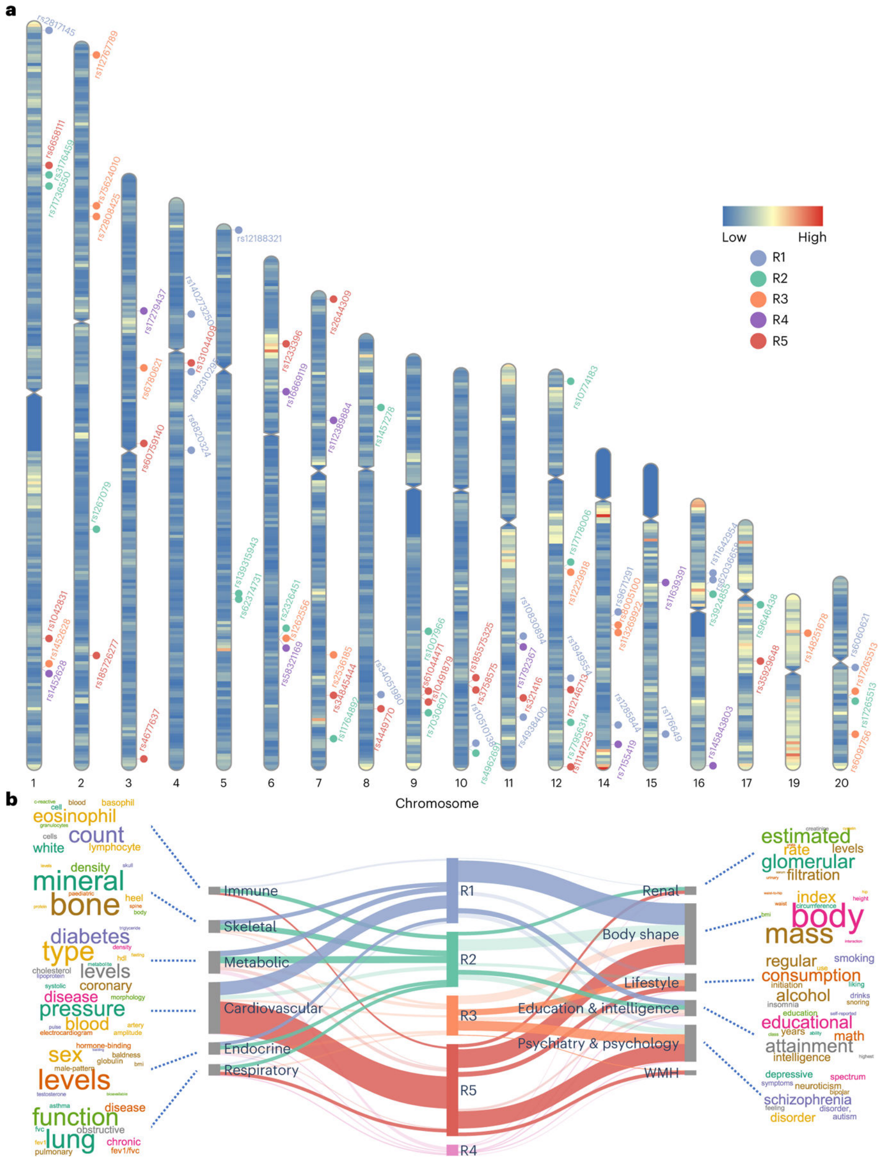


Fig. 5 |. Five R-indices were associated with 73 genomic loci.

a, Overall, 73 genomic loci were associated with the five R-indices using a genome-wide *P* value threshold ($-\log_{10}(P) > 7.30$). For visualization purposes, we annotated the locus with the top lead SNP. Two-sided *t*-tests were applied for testing the significance of regression coefficients of SNPs. **b**, Phenome-wide associations of our identified genomic loci in the EMBL-EBI GWAS catalog (query date, 2 July 2023 via FUMA103 v.1.5.4). We examined the candidate and independent significant SNPs within each genomic locus and connected them to various clinical traits through a comprehensive query. The width of each connection represented the number of associations between the genomic loci revealed in our study and clinical traits in the literature. These traits were grouped into high-level categories, including different organ systems, psychiatric and psychological conditions and lifestyle factors, body shape, etc. To enhance visual understanding of each category, we generated keyword cloud plots based on the frequency of clinical traits within each category. We excluded brain structure-related traits, which were expected to have the highest number of associations with the SNPs we identified.

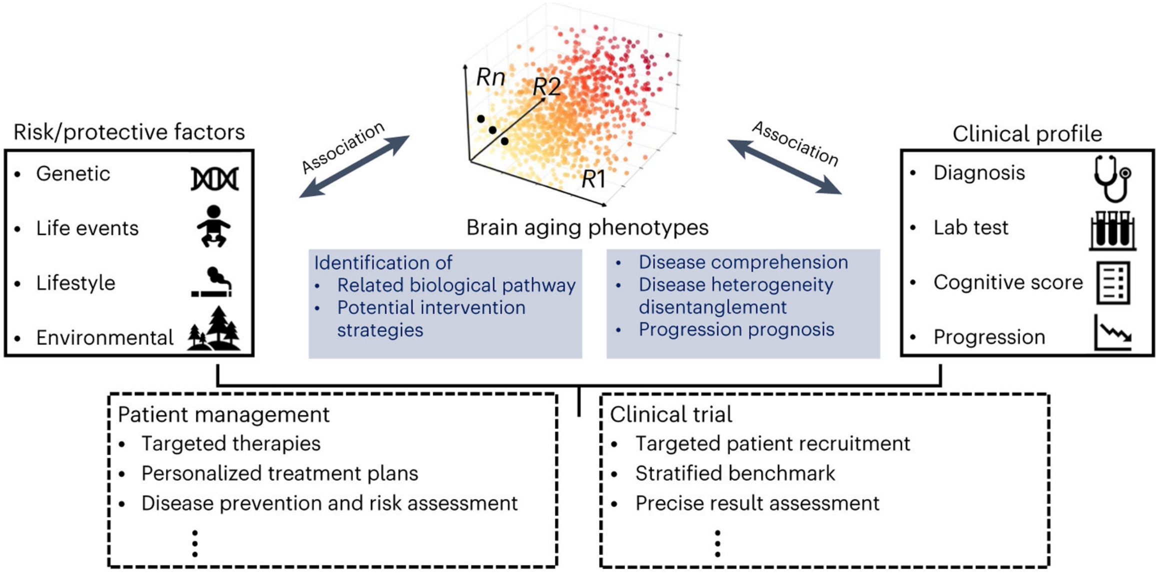


Fig. 6 |. The R-indices can have broad implications for healthcare.

The deep-learning-derived R-indices are derived brain aging phenotypes that can serve as endophenotypes, or intermediate phenotypes, of diverse underlying neuropathologic processes that accompany aging. They also aid in understanding the risk and protective factors contributing to this heterogeneity. More notably, these R-indices, combined with risk factors and clinical profiles, establish a concrete system for personalized patient management and targeted clinical trial recruitment designs.

Table 1 |

Participants and studies for model training and data analyses

Study	Sample size		Age	Sex	MCI/dementia	Training sample		Longitudinal data	Mean follow-up years
	Pre-aging	Aging				Pre-aging	Aging		
ADNI	0	2,436	73.1 ± 7.3	52.3%	1,101/419	0	1,000	1,911	3.6 ± 2.9
UKBB	900	38,675	64.1 ± 9.4	47.1%	0/1	300	1,000	1,382	2.3 ± 0.1
BLSA	161	956	65.6 ± 10.4	47.4%	11/8	161	956	694	6.8 ± 4.7
AIBL	5	968	73.0 ± 7.1	44.0%	130/92	5	968	459	4.1 ± 2.4
BIOCARD	37	270	58.4 ± 8.5	41.7%	3/1	37	270	235	13.3 ± 6.5
OASIS	35	1,053	70.3 ± 11.5	44.4%	0/239	35	1,000	518	4.9 ± 2.8
PENN	28	1,100	72.6 ± 11.4	42.4%	270/277	28	1,000	200	2.8 ± 2.4
WRAP	7	264	61.7 ± 9.3	28.4%	0/0	7	264	180	5.7 ± 2.5
CARDIA	280	532	51.3 ± 11.8	46.7%	0/0	280	532	270	4.7 ± 0.4
SHIP	1,384	1,810	52.8 ± 14.2	48.1%	0/0	300	1,000	-	-
WHIMS	0	1,418	69.6 ± 9.3	0%	0/0	0	1,000	727	4.7 ± 0.4

For age and follow-up years, the mean and s.d. are reported. For sex, the percentage of males is presented. Longitudinal data, number of participants with longitudinal data available. MCI/dementia indicates the number of participants with these two diagnoses at baseline. BLSA, Baltimore Longitudinal Study of Aging; AIBL, Australian Imaging, Biomarker, and Lifestyle study of aging; BIOCARD, Biomarkers of Cognitive Decline Among Normal Individuals in the John Hopkins University study; OASIS, Open Access Series of Imaging Studies; PENN, University of Pennsylvania Memory Center cohort; WRAP, Wisconsin Registry for Alzheimer’s Prevention; CARDIA, Coronary Artery Risk Development in Young Adults; SHIP, Study of Health in Pomerania; WHIMS, Women’s Health Initiative Memory Study; -, no longitudinal data available.