

ARTICLE OPEN



Neurofind: using deep learning to make individualised inferences in brain-based disorders

S. Vieira^{1,2,3}, L. Baecker¹, W. H. L. Pinaya^{1,4}, R. Garcia-Dias¹, C. Scarpazza^{1,5,6}, V. Calhoun^{1,7} and A. Mechelli¹✉

© The Author(s) 2025

Within precision psychiatry, there is a growing interest in normative models given their ability to parse heterogeneity. While they are intuitive and informative, the technical expertise and resources required to develop normative models may not be accessible to most researchers. Here we present Neurofind, a new freely available tool that bridges this gap by wrapping sound and previously tested methods on data harmonisation and advanced normative models into a web-based platform that requires minimal input from the user. We explain how Neurofind was developed, how to use the Neurofind website in four simple steps (www.neurofind.ai), and provide exemplar applications. Neurofind takes as input structural MRI images and outputs two main metrics derived from independent normative models: (1) Outlier Index Score, a deviation score from the normative brain morphology, and (2) Brain Age, the predicted age based on an individual's brain morphometry. The tool was trained on 3362 images of healthy controls aged 20–80 from publicly available datasets. The volume of 101 cortical and subcortical regions was extracted and modelled with an adversarial autoencoder for the Outlier index model and a support vector regression for the Brain age model. To illustrate potential applications, we applied Neurofind to 364 images from three independent datasets of patients diagnosed with Alzheimer's disease and schizophrenia. In Alzheimer's disease, 55.2% of patients had very extreme Outlier Index Scores, mostly driven by larger deviations in temporal-limbic structures and ventricles. Patients were also homogeneous in how they deviated from the norm. Conversely, only 30.1% of schizophrenia patients were extreme outliers, due to deviations in the hippocampus and pallidum, and patients tended to be more heterogeneous than controls. Both groups showed signs of accelerated brain ageing.

Translational Psychiatry (2025)15:69; <https://doi.org/10.1038/s41398-025-03290-x>

INTRODUCTION

Psychiatric and neurological disorders represent 10.4% of the global burden of disease and account for 258 million DALYs per year [1]. Until recently, the investigation of the brain correlates of these disorders has relied mostly on analytical methods based on mass-univariate comparisons between groups (e.g., statistical parametric mapping) [2–4]. However, this approach is not aligned with the complex, widespread and heterogeneous alterations which are typical of these disorders and the need for individual-level decisions in clinical practice [2, 5, 6]. This discrepancy may explain the limited translational impact of neuroimaging findings in neurology and psychiatry so far [7]. In recent years, a growing number of studies have attempted to address this issue by using multivariate predictive methods that allow statistical inferences at the individual level [8–11]. Within this movement, there is a growing interest in normative models given their added ability to parse heterogeneity [12–14]. Here, the aim is to first model the variability of a measure (e.g., brain morphology) or relationship (e.g., brain morphology and age) of interest in a reference cohort. Because the latter is typically a group of healthy controls (HC), it is

assumed that the variability captured by the model corresponds to the 'norm', i.e., expected pattern in the absence of illness. A new and unseen individual can then be mapped against this model, and the distance of this individual from the norm can be quantified. Within this approach, individuals with the same diagnosis can differ from healthy controls in their own unique way or, importantly, not differ at all (i.e., not deviate from the norm) [15]. This is a significant departure from the traditional case-control design, where the 'average patient' is compared with the 'average control', under the assumption that they represent qualitatively different groups. This may be a reasonable assumption when the patient group is characterised by consistent and well-defined alterations. However, a growing number of studies using normative modelling are revealing that a surprisingly large proportion of patients fall within the normative range for several brain features [16, 17] and that there is substantial heterogeneity within diagnoses [6, 16, 18–22] as well as overlap between diagnoses [18].

Following these advances and potential for clinical translation, several web-based tools have been developed. However, as we

¹Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. ²Department of Radiology, Lausanne University Hospital and University of Lausanne (CHUV-UNIL), Lausanne, Switzerland. ³Center for Research in Neuropsychology and Cognitive Behavioural Intervention, Faculty of Psychology and Educational Sciences, University of Coimbra, Coimbra, Portugal. ⁴Department of Biomedical Engineering, King's College London, London, UK. ⁵Department of General Psychology, University of Padova, Padova, Italy. ⁶IRCCS S Camillo Hospital, Venezia, Italy. ⁷Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS) [Georgia State University, Georgia Institute of Technology, and Emory University], Atlanta, GA, USA. ✉email: a.mechelli@kcl.ac.uk

Received: 13 March 2024 Revised: 17 January 2025 Accepted: 18 February 2025

Published online: 27 February 2025

discuss in a recent systematic review [23], these tools were designed to support the diagnosis of neurological disorders namely dementia (ADABOOST [24], Jung Diagnostics [25, 26], NeuroQuant [27], Quantib [28], volBrain [29, 30]), multiple sclerosis (Jung Diagnostics [31], Quantib [32], volBrain [33]), traumatic brain injury (Icbrain [34], NeuroQuant [35–38]), temporal lobe epilepsy (NeuroQuant [39, 40]) and intracranial hemorrhages and mass effects in the brain (Qure [41]). These tools map known biomarkers for these disorders including atrophy of hippocampal/sub-cortical structures and ventricles (NeuroQuant, Jung Diagnostics), white-matter lesions (Icbrain, Jung Diagnostics, Quantib, volBrain) or gross abnormalities with computed tomography (Qure). Therefore, to our knowledge, there are no easy-access tools to map deviations from normative whole-brain morphology without making assumptions about a possible diagnosis of a neurological disorder. This hinders the use of existing tools in psychiatry, where disorders are characterised by a heterogeneous, subtle and widespread pattern of abnormalities across the brain. Here we present *Neurofind*, a new user-friendly and freely available web-based tool that outputs an individualised report based on normative modelling from high-resolution structural magnetic resonance (MRI) images. This individualised report includes two main components: (1) Outlier Detection - overall deviation from the normative brain morphology as well as deviations for each brain region, and (2) Brain Age - predicted age based on an individual's brain morphology. The Outlier Detection model uses an adversarial auto-encoder (AAE), a deep learning approach that has shown promising results in clinical neuroimaging [42, 43]. Specifically, we developed an AAE that learns a latent representation of the input data by first reducing its dimensionality via consecutive layers of non-linear transformations (encoder); this reduced representation is then used to reconstruct the input (decoder). By constraining this reconstruction to be as similar as possible to the input data, the autoencoder is forced to learn a good representation of the input data [44]. Therefore, the AAE will be able to reconstruct new data with minimal difference between input and reconstruction if it is similar to the data used to learn the latent representation. In *Neurofind*, we used an AAE to learn a latent representation of the normative brain morphology. It is expected that, when presented with new images, the AAE will be able to reconstruct data from healthy controls, but it will be less precise when processing data from patients, i.e., the reconstruction error (i.e., difference between input and reconstruction) will be larger in patients compared to healthy controls. Therefore, the reconstruction error can be thought of as a proxy for deviation from the normative disease-free brain morphology. The Brain Age model, on the other hand, measures the effects of ageing on the brain and builds on the well-established relationship between age and neuroanatomy across the lifespan [45]. Brain age is often used to calculate the brain age gap (BAG), i.e., the difference between the predicted brain age and chronological age. A positive BAG means that the individual's predicted brain age was higher than their actual age, suggesting 'accelerated' ageing; conversely, a negative BAG indicates 'delayed' ageing. There has been a surge in the investigation of BAG as a potential biomarker [46] after several studies showing an higher brain ages in psychosis [47–51], bipolar disorder [50–53], Alzheimer's disease (AD) [51, 54, 55], among others. In addition to providing these two informative and complementary metrics, *Neurofind* uses a novel method – Neuroharmony [56] – to mitigate scanner effects of new incoming images. This overcomes the limitation of popular harmonisation methods, such as ComBat [57, 58], that require a representative sample from each scanner, which is not suitable for clinical translation, as new patients' scans will likely originate from previously unseen scanners. Therefore, *Neurofind* combines our previous work on the use of AAE to build a normative model of brain morphology [59], brain age [60] and harmonisation of unseen scanners [56] into a user-friendly web-based tool targeted

at researchers who wish to apply normative modelling to their morphological data. In the following sections, we present (1) the development of the models in *Neurofind*, (2) a guide on how to use *Neurofind* in four simple steps, and (3) exemplar applications of the tool in AD and schizophrenia (SZ).

MODEL DEVELOPMENT

Datasets

Data for training the web-based normative models used in *Neurofind* consisted of four publicly available datasets of healthy controls, namely the Human Connectome Project [61, 62], the Human Connectome Project Aging [63], Biobank [64] and IXI (<http://brain-development.org/ixi-dataset/>).

MRI data acquisition and pre-processing

The initial total sample included 13,918 T1-weighted MRI images from 7 scanners (see Supplementary Materials for image acquisition parameters). Participants with missing data and younger than 20 and older than 80 years old were excluded. Poor quality images were excluded based on the MRIQC tool [65]. This tool uses 68 image quality metrics (IQMs) such as the presence of movement, artefacts, and signal-to-noise ratio to determine the probability of an image being unusable. Images with a probability higher than 0.7 were discarded. The remaining images were pre-processed using the recon-all pipeline with standard parameters in FreeSurfer (version 6.0) [66]. The cortical surface was parcellated using the Desikan-Killiany cortical atlas [67] and segmented into 68 cortical regions (34 per hemisphere). An additional 33 neuroanatomical structures were extracted using the ASEG atlas in FreeSurfer [66, 67] totalling 101 brain regions (ROIs) (full list of extracted regions in the Supplementary Materials). We estimated the relative brain region volumes (ROI_{rel}) for each subject by dividing the original brain region volumes by the total intracranial volume (also computed with FreeSurfer). The final pre-processed data included 13,187 participants. However, there was an imbalanced distribution between the number of younger and older healthy controls, mostly due to the Biobank dataset (Fig. 1). The largest scanner of the Biobank dataset was under-sampled such that a maximum of 55 participants per age group were selected at random. Since outliers are unexpected in healthy subjects and are likely to be due to artefacts, healthy controls in the train set with at least 10 ROI_{rel} more than 3 standard deviations (σ) away from the sample mean (μ) were considered outliers and excluded. This process was repeated iteratively, recalculating μ and σ until no additional participant met the criteria for being an outlier. This process was implemented within each scanner to ensure that participants would not be considered outliers simply due to differences between scanners. The final data for the train set comprised 3362 participants (Fig. 1, Table 1, Supplementary Materials for sample size/scanner).

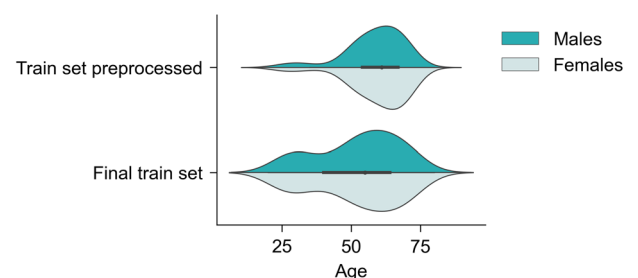


Fig. 1 Distribution of age and sex in the train set before and after addressing data imbalance.

Table 1. Demographic characteristics of the train set used in model development.

	Train set (N = 3362)
Sex F/M, N (%)	1893 (56.3) / 1469 (43.7)
Age, median [Q1, Q3]	52.0 [15.2]

Scanner harmonisation

Scanner effects on the ROI_{rel} were mitigated using the Neuroharmony method [56]. Briefly, Neuroharmony uses a random forest model to capture the relationship between IQMs and the corrections for each brain region prescribed by ComBat, a popular harmonisation approach that uses an empirical Bayes framework to correct for additive and multiplicative batch effects [68]. Once this relationship is learned, Neuroharmony predicts and applies the ComBat corrections for a new set of IQMs extracted from an unseen T1-weighted image. Neuroharmony was trained twice to accommodate the specific needs of the Outlier Detection and Brain Age models. For the former, Neuroharmony was trained whilst controlling for the effects of age and sex, whereas for Brain Age, only sex was added as a covariate. Neuroharmony was trained in the same train datasets as used for the development of the outlier detection and brain age models. Results of the harmonisation are shown in the Supplementary Materials.

Outlier detection

Our outlier detection model uses an adversarial autoencoder (AAE) to learn a representation of disease-free brain morphology. Briefly, AAE combines a standard autoencoder with adversarial learning to improve the learned latent representation by forcing it to have a distribution similar to a desired prior distribution. In our approach, we used a Gaussian distribution. This is achieved by adding a third element to the standard autoencoder, the discriminator, responsible for deciding whether its input data comes from random numbers sampled from the predefined prior distribution or the latent code. A successful AAE will be able to generate a latent code capable of fooling the discriminator into inferring that the encoded samples come from the prior distribution (i.e., the discriminator will not be able to distinguish the two distributions). Therefore, the aim is to produce a latent representation that yields a low reconstruction error whilst also having a similar shape to the desired prior distribution. The AAE was trained in the harmonised ROI_{rel} from the train set following the same approach described in previous work [59] (see Supplementary Materials for details on model training). The reconstruction error for each ROI_{rel} was estimated by calculating the squared error between the inputted value and its reconstruction. An overall reconstruction error was then generated for each participant by calculating the mean squared error between the reconstruction and the inputted data. The distributions for all ROI -level and overall reconstruction errors in the train set were positively skewed. Therefore, all values were scaled by subtracting the median and dividing by the interquartile range such that the median of the overall and ROI -level scaled errors was 0.0 (IQR = 1.0) (Supplementary Materials). The final scaled errors, which we named Outlier Index Score (OIS), represent the number of interquartile ranges away from the median in the train set. The OIS is categorised into: Within the norm (<0.26), Low (≥ 0.26), Medium (≥ 1.1) and High (≥ 2.3) deviations from the norm (values indicate number of interquartile ranges away from the median in the train set). Cut-off values were extracted from the train set by first dividing the distribution of OIS into centiles and then grouping them into quartiles.

Brain age

To predict brain age, harmonised and scaled ROI_{rel} from the train set were inputted to a linear support vector regression (SVR)

model as implemented in the Python package scikit-learn [69] with a similar approach as described previously [60]. A systematic hyperparameter search for C was conducted using a ten-fold cross-validated grid-search over the search space 2^{-7} , 2^{-5} , 2^{-3} , 2^{-2} , 2^{-1} , 2^0 , and 2^1 . The C value with the best performance, as determined by the scoring parameter negative mean absolute error (MAE), was then applied to the whole training set ($C = 2^{-2}$). The parameters epsilon and tolerance for stopping criterion were 0.1 and $1e-3$, respectively. In the train set, the MAE was 10.5 ± 9.8 years.

HOW TO USE NEUROFIND

Neurofind is a web-based research tool available at www.neurofind.ai. This section provides a step-by-step guide on how to get started, upload your images, check the status of your analyses as well as download and interpret the results. A graphical illustration of these steps can be found in Fig. 2. Details regarding how the data are stored and managed can be found in the Privacy Policy: <https://neurofind.ai/privacy-policy>.

Step 1. Registration

To use Neurofind, you will first need to register a user account. You will be asked to provide simple key information (e.g., name, institution, role) and to read and accept the privacy policy. You will receive an email confirmation that your registration is approved.

Step 2. Upload image(s)

Neurofind allows the processing of a single image or several images in bulk. To upload one image, choose Single upload from the top menu and fill in the necessary information. Upload your image by dragging it or clicking on Select files and navigating to the folder where the image is stored. Images can only be in NIfTI (.nii, .nii.gz) format. It is important that you have the subject's consent to analyse their data in Neurofind, so the upload page also asks you to confirm that you have the necessary approvals before you can select Submit. To upload several images at once, select Multiple upload from the top menu. You will be asked to upload all your images at once along with a CSV file with all the relevant information about the images. To do this, select Download CSV to view the instructions on how to fill in the CSV file and download the template. It is important that you do not include any identifiable information in the CSV file or elsewhere on the website. Once completed, upload it by pressing Upload CSV. Next, click on Upload scans to upload the images. The processing of the image(s) will start as soon as the upload is completed.

Step 3. Check progress

To monitor the progress of your analyses, go to My scans in the top menu. You will see the status of all images submitted for processing as well as ID, key demographic information, and time when processing began. The status of each image will be one of the following: Missing information, Processing, Completed or Failed. If any of the information requested in Step 2 is missing, processing will not begin automatically, and the status will be Missing information. To add missing information, go to My scans, find the image with the status Missing information and select Info. Once the missing information has been entered, processing will begin automatically, and the status will change to Processing. Processing includes the entire analysis, from image upload to results. This means that the pre-processing of the images (e.g., segmentation into different brain volumes) is also done automatically by Neurofind. If there is a problem during processing, the status will be changed to Failed. Possible reasons for this include poor-quality images or issues with network connection. If this occurs, you will need to reupload the image(s) from the beginning. When Processing is finished, the status will change to

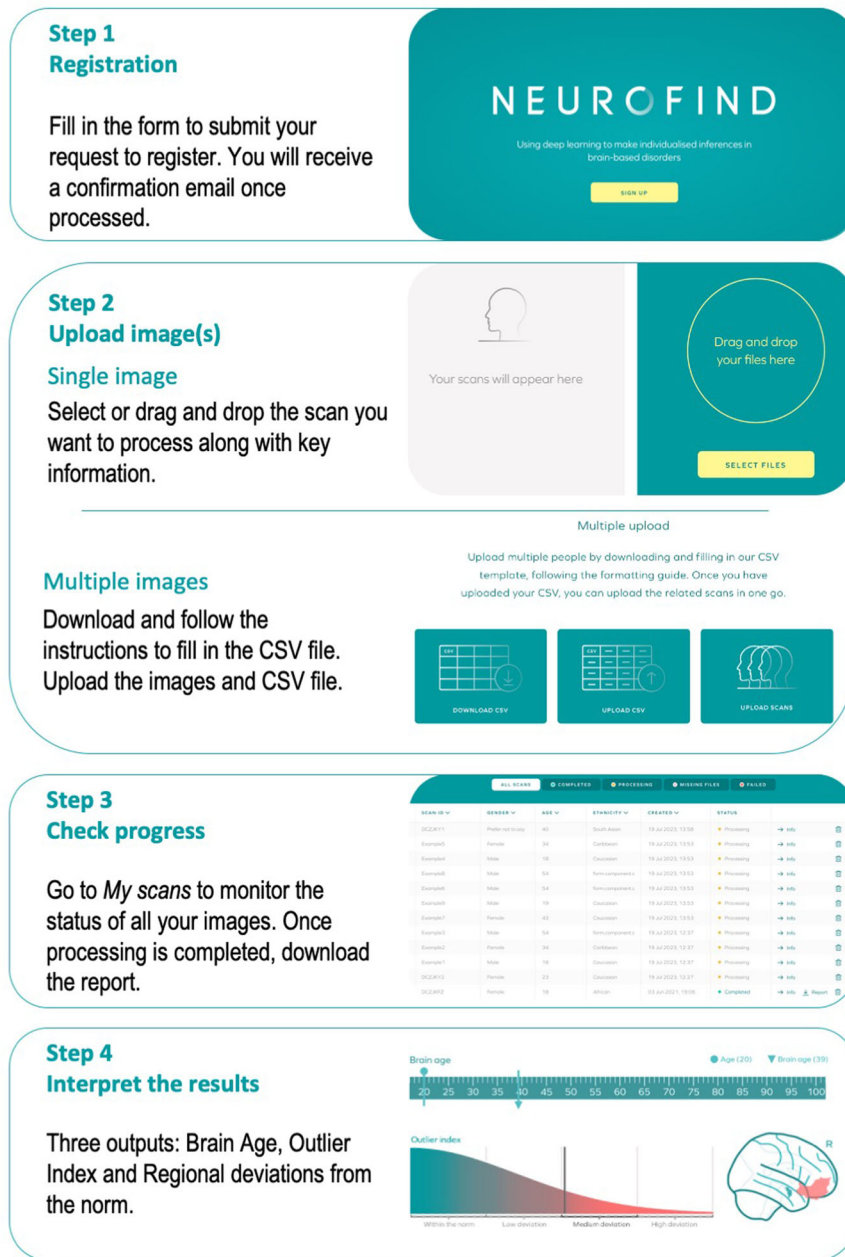


Fig. 2 Main steps to use Neurofind.

Completed and an individualised report will be available to view and download as a PDF file. You will also be notified via email once processing is complete.

Step 4. Obtaining the results

The individualised report contains three main outputs: Brain Age, Outlier Index Score and Regional deviations from the norm. For Brain Age, the predicted and chronological ages are given. The difference between the two represents the BAG. A positive BAG means that the individual's predicted brain age was higher than their chronological age, suggesting 'accelerated' ageing; conversely, a negative BAG indicates 'delayed' ageing. The Outlier Index Score conveys the magnitude of the overall deviation from the normative brain morphology. As the deviation increases, the Outlier Index Score will be classified as Within the norm, Low, Medium and High deviation from the norm (see Section 'Outlier detection' for explanation of cut-off values). The brain regions with

the largest deviations from the norm are shown in a glass brain and shown in full in a table at the end of the report.

EXEMPLAR APPLICATIONS

Datasets used for exemplar applications

Exemplar applications were carried out in three clinical datasets: the Australian Imaging, Biomarkers and Lifestyle Study (AIBL), the MIND Clinical Imaging Consortium (MCIC) and the Center for Biomedical Research Excellence (COBRE) dataset. The AIBL dataset includes patients diagnosed with AD and healthy controls; the latter two datasets include patients diagnosed with SZ and corresponding healthy controls; for details about recruitment see [70–73].

MRI data acquisition and pre-processing

The initial total sample for exemplar applications included 759 T1-weighted MRI images from 10 scanners (see Supplementary

Materials for image acquisition parameters). Data were cleaned and pre-processed in line with the steps outlined for the model development data in Section 'Model development'. In addition, healthy controls from AIBL were randomly subsampled to match the sample size of AD patients to address the large class imbalance. The final data comprised 364 participants in the clinical datasets (Fig. 3, Table 2, Supplementary Materials for sample size/scanner). Age did not follow a normal distribution in all sub-groups according to the Shapiro-Wilk test ($AD_{AD} = 0.9$, $p < 0.001$; $AD_{HC} = 1.0$, $p < 0.05$; $SZ_{SZ} = 0.9$, $p < 0.001$; $SZ_{HC} = 0.9$, $p < 0.01$). Differences in sex and age between patients and healthy controls in the clinical data were tested using chi-square test and Mann-Whitney U test, respectively; this revealed no statistically significant differences between the two groups (Table 2).

Scanner harmonisation

The Neuroharmony model (trained in the model development data, as described in Section 'Model development') was applied separately to the clinical datasets used for exemplar applications. Results of the harmonisation are shown in the Supplementary Materials.

Statistical analyses in exemplar applications

The overall and ROI-level reconstruction errors in the clinical sample were scaled according to the median and interquartile range obtained in the train set (see Section 'Outlier detection'). Differences in overall and ROI-level OIS and BAG between patients and respective controls were investigated using an independent-sample t-test or Mann-Whitney U Test depending on whether the distribution for both groups was normally distributed or not, respectively. Normality was tested using the Shapiro-Wilk test. For the ROI-level comparisons the p-value was adjusted using the Benjamini-Hochberg correction for multiple comparisons. The association between OIS and BAG with symptom severity was tested with Pearson's correlation coefficient (r) or Spearman's rho (ρ) depending on the outcome of the Shapiro-Wilk test for the symptoms. Statistical significance was set at 0.05. Patients' symptom severity was assessed with the Mini Mental State Examination (MMSE) [74] (AIBL), the Scale for the Assessment of Negative Symptoms (SANS) [75] and the Scale for the Assessment of Positive Symptoms (SAPS) [76] (MCIC) or the Positive and Negative Syndrome Scale (PANSS) [77] (COBRE). Lower scores in

the MMSE and higher scores in the psychosis scales indicate worse illness severity. The positive (PANSS Positive and SAPS) and negative (PANSS Negative and SANS) symptoms scales were normalised to ensure comparability across sites using the formula [4]:

$$\text{New score} = \frac{\text{Individual raw score} - \text{Minimum}}{\text{Maximum} - \text{Minimum}}$$

where Minimum and Maximum refer to the lowest and highest score allowed for either PANSS or SAPS/SANS. The resulting symptom severity scores were normalised by subtracting the mean from every item and then dividing the resulting value by the standard deviation of the item (i.e., zero mean unit variance normalization). To investigate heterogeneity of deviations from the normative brain morphology amongst participants, we calculated the mean pairwise cosine similarity (CS) across all individual ROI-level OIS and within predefined brain regions based on the Desikan-Killiany atlas: frontal, temporal, parietal and occipital lobes, insula, cingulate, subcortical structures, cerebellum, corpus callosum and ventricles (Supplementary Materials). Briefly, CS quantifies the similarity between two vectors by calculating the cosine of the angle between the two. The magnitude varies from 0 (low similarity between individuals, i.e., vectors are 90 degrees or perpendicular to each other) and 1 (perfect similarity between individuals). BAG was calculated by subtracting the chronological age from the predicted brain age and normalised according to the train set used for model development.

Results from exemplar applications

Quantifying deviation from the normative brain morphology. The overall and ROI-level OIS did not follow a normal distribution, whilst the BAG did (Supplementary Materials). AD patients (Median = 2.46, IQR = 2.50) showed a significantly larger OIS compared to HC (Median = 0.91, IQR = 1.36) (Mann-U = 1361; $p < 0.001$) (Fig. 4A). AD patients showed significantly larger median OIS for individual ROIs relative to respective HC for bilateral amygdala (Mann-U_{Left Amyg} = 1196, $p < 0.01$; Mann-U_{Right Amyg} = 1246, $p < 0.001$), hippocampus (Mann-U_{Left Hipp} = 1476, $p < 0.001$; Mann-U_{Right Hipp} = 1425, $p < 0.001$), inferior temporal gyrus (Mann-U_{Left Inf Temp} = 1245, $p < 0.001$; Mann-U_{Right Inf Temp} = 1343, $p < 0.001$) and inferior lateral ventricle (Mann-U_{Left Inf. Lat. Ven.} = 1519, $p < 0.001$; Mann-U_{Right Inf. Lat. Ven.} = 1382, $p < 0.001$), left inferior parietal gyrus (Mann-U = 1185, $p < 0.05$) (Fig. 4C; see Supplementary Materials for detailed results). Conversely, there was no statistically significant difference in overall OIS between SZ patients (Median = 1.38, IQR = 1.83) and respective HC (Median = 1.10, IQR = 1.76) (Mann-U = 7657; $p = 0.091$) (Fig. 4A). However, patients showed significantly larger median deviations from HC in bilateral hippocampus (Mann-U_{Left Hipp} = 8077, $p < 0.01$; Mann-U_{Right Hipp} = 7384, $p < 0.05$) and left putamen (Mann-U = 7846, $p < 0.05$) (Fig. 4D, see Supplementary Materials for detailed results).

Only 3.4% of AD patients against 25.9% HC had an overall OIS within the norm (Fig. 5). The proportion of AD patients with low, medium and high deviation relative to HC increased steadily such that 55.2% of patients and only 13.8% HC had high deviations

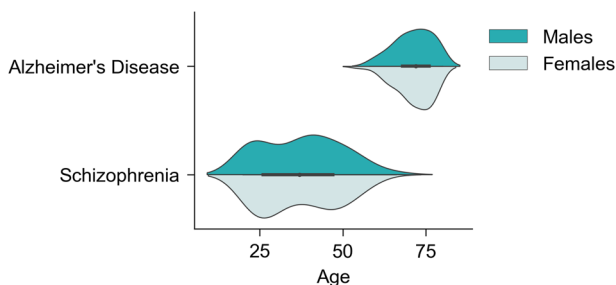


Fig. 3 Distribution of age and sex in the clinical datasets used for exemplar applications.

Table 2. Demographic characteristics of the clinical datasets used for exemplar applications.

	Clinical datasets (N = 367)					
	Alzheimer's disease (N = 122)			Schizophrenia (N = 245)		
	AD (N = 61)	HC (N = 61)	p	SZ (N = 134)	HC (N = 111)	p
Sex F/M, N (%)	37 (60.7) / 24 (39.3)	43 (70.5) / 18 (29.5)	0.341	30 (22.4) / 104 (77.6)	32 (28.8) / 79 (71.2)	0.314
Age, median [Q1,Q3]	73.0 [69.0,77.0]	71.0 [68.0,75.0]	0.244	35.0 [25.0,46.0]	37.0 [27.0,47.0]	0.282

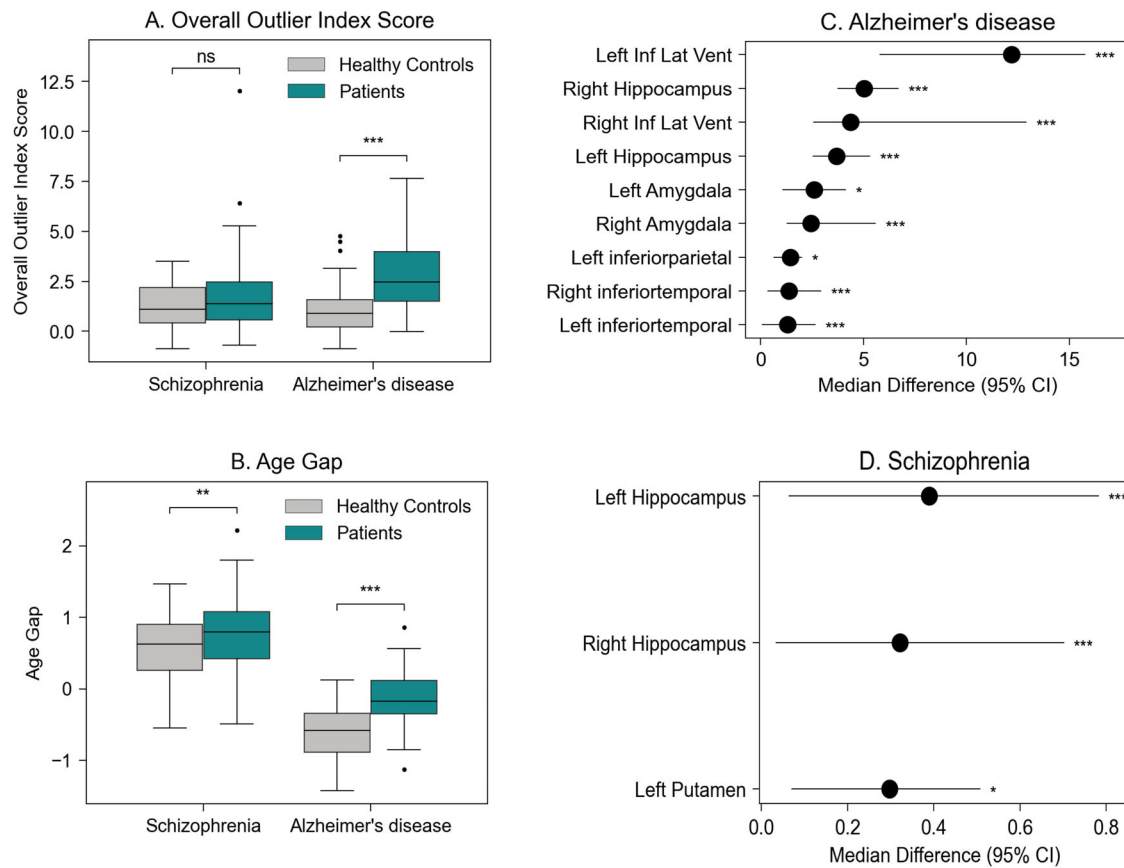


Fig. 4 Results from the exemplar applications for AD and SZ. **A** Outlier Index Score (OIS) for AD and SZ. **B** BAG for patients and HC. **C, D** Median difference between OIS for schizophrenia / Alzheimer's disease and respective controls for ROIs with a statistically significant difference between the two. Plots show the mean difference between patients and respective controls and the 95% confidence interval (CI) for this difference. The 95% CI was calculated with bootstrapping (1000 repetitions). *** $p \leq 0.001$, ** $p \leq 0.010$, * $p < 0.05$, ns $p > 0.05$.

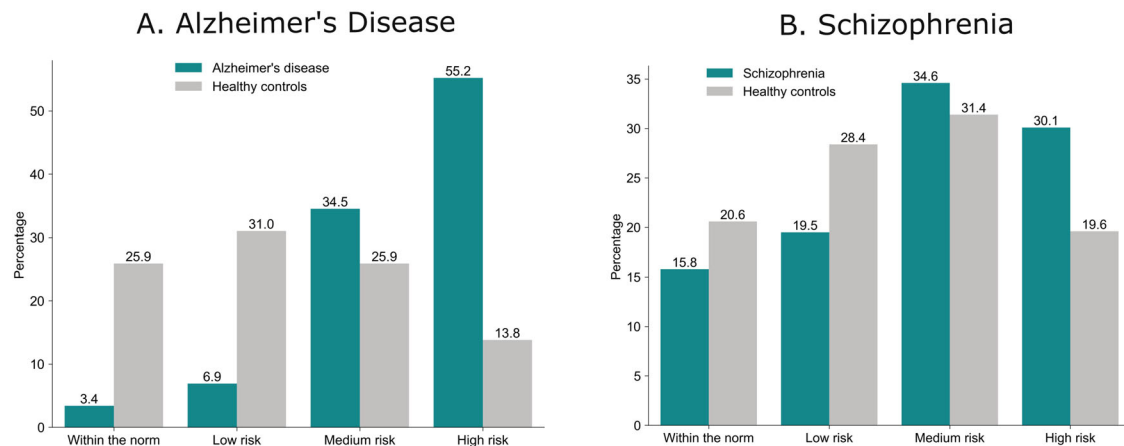


Fig. 5 Proportion of patients with overall OIS within the norm, low, medium, or high deviation. **A** In the Alzheimer's disease group, only a small fraction of patients exhibited an overall OIS within the normative range. The percentage of patients classified with deviations increased progressively, with the majority falling into the high deviation category. **B** In the schizophrenia group, a greater proportion of patients displayed deviations from the norm, with more individuals classified in the medium and high deviation categories than in the low deviation range.

from the norm. Conversely, 15.8% of SZ patients and 20.6% of HC had an overall normative brain morphology. However, there was higher proportion of SZ patients in the medium and high deviation categories compared to HC.

Estimating brain age gap. The overall performance of the brain age model in the healthy controls of the clinical datasets was MAE

10.0 ± 6.3 years. The mean scaled BAG was significantly larger for AD ($M = -0.1$, $SD = 0.4$) compared to their respective HC ($M = -0.6$, $SD = 0.4$) ($p < 0.001$) as well as for SZ ($M = 0.8$, $SD = 0.5$) compared to their respective HC ($M = 0.6$, $SD = 0.4$; $p < 0.01$). This indicates that for both illnesses, the predicted brain age was higher than their expected age based on the control groups, suggesting 'accelerated' ageing (Fig. 4D).

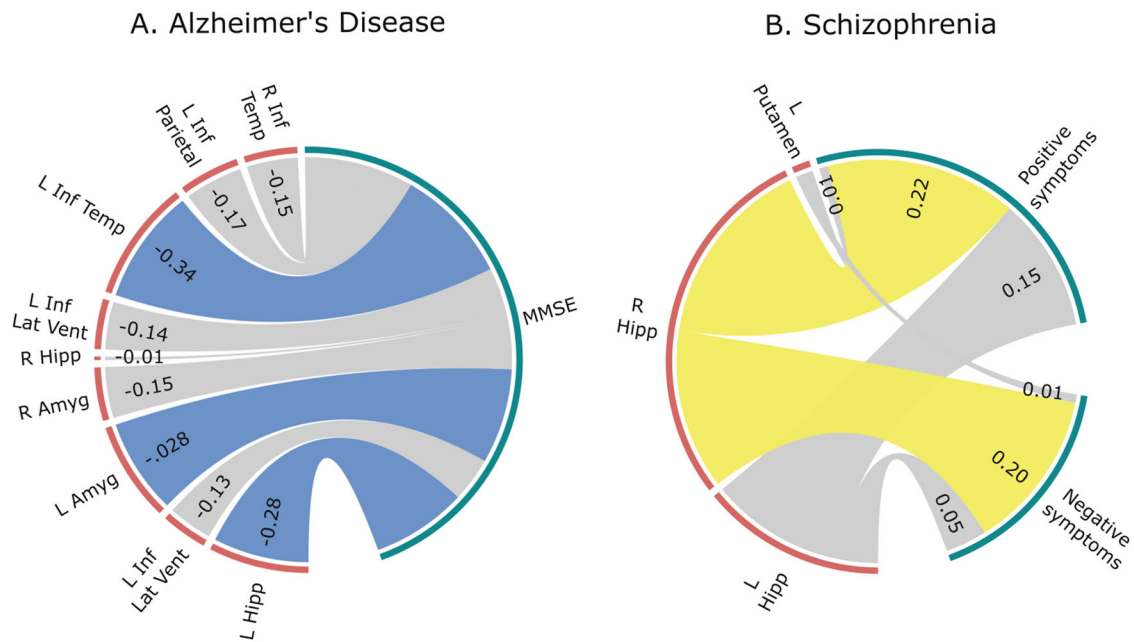


Fig. 6 Spearman's correlation between regions of interest (ROIs) with significantly larger deviations from the norm in patients and symptom severity. **A** In Alzheimer's disease, greater deviations from the norm in the left hippocampus, left amygdala, and left inferior temporal gyrus were associated with increased disease severity, as indicated by significant negative correlations. **B** In schizophrenia, larger deviations in the right hippocampus were significantly associated with greater positive and negative symptom severity.

Investigating the neural correlates of symptom severity. Severity of AD was negatively associated with the left hippocampus ($\rho = -0.28$, $p < 0.05$), left amygdala ($\rho = -0.28$, $p < 0.05$) and left inferior temporal gyrus ($\rho = -0.34$, $p < 0.05$). As expected, these negative associations indicate that the larger the deviation from the norm, the more severe the symptoms. In SZ, there was a significant positive association between right hippocampus and positive ($\rho = 0.22$, $p < 0.05$) and negative ($\rho = 0.20$, $p < 0.05$) symptoms (Fig. 6, see Supplementary Materials for detailed results). The association between BAG and symptom severity was statistically significant in SZ for positive ($\rho = 0.20$, $p < 0.05$) but not negative symptoms ($\rho = 0.04$, $p = 0.680$) nor MMSE in AD ($\rho = -0.16$, $p = 0.266$).

Estimating heterogeneity. Across all ROIs, AD patients were more homogeneous (Mean CS = 0.54, SD = 0.22) than their corresponding HC (Mean CS = 0.31, SD = 0.25). Higher homogeneity in patients was found for the ventricles (Mean CS_{HC} = 0.56, SD = 0.24; Mean CS_{AD} = 0.72, SD = 0.20) and subcortical regions (Mean CS_{HC} = 0.42, SD = 0.20; Mean CS_{AD} = 0.60, SD = 0.20). Conversely, SZ patients were slightly more heterogeneous (Mean CS = 0.32, SD = 0.15) than HC (Mean CS = 0.40, SD = 0.17) across all ROIs. This was most visible in the ventricles (CS_{HC} = 0.62, SD = 0.23; CS_{SZ} = 0.55, SD = 0.22) and subcortical regions (CS_{HC} = 0.47, SD = 0.21; CS_{SZ} = 0.38, SD = 0.20) (Fig. 7, see Supplementary Materials for detailed results).

DISCUSSION

The recent widespread interest in personalised psychiatry and neurology, combined with the increasing availability of large imaging datasets, has propelled a renewed interest in normative modelling [12, 15, 78]. This approach involves first modelling a feature of interest in a reference cohort, usually healthy controls, and then mapping patients against this model. It is then possible to calculate the deviation from the reference cohort, also known as the 'norm', for each individual patient. Interesting findings include evidence that a large proportion of psychiatric and

neurological patients fall within the normative range for brain morphology [16, 17], and that there is high degree of heterogeneity within a diagnosis [6, 18, 19] and overlap between diagnoses [18]. These results are consistent with the current understanding of mental health disorders as dimensional constructs [79], and challenge the premise of the widely used case-control design.

An important next step is to further explore 'deviation from the norm' as a potential biomarker. This might involve, for example, estimating deviation from the norm in several clinical populations, investigating what drives extreme deviations within a diagnosis and how such deviations relate to clinical outcomes or other relevant variables such as genetic risk and cognitive performance. On the one hand, the typical output of a normative model is intuitive, informative and easy to use in further statistical analyses. On the other hand, the development of normative models may not be feasible for most researchers due to lack of technical expertise, adequate datasets or/and computational resources. Neurofind was developed to bridge this gap and make normative modelling accessible to the wider research community. A number of strategies were adopted to maximise the usability and utility of the tool [80]. Firstly, Neurofind was developed as a *user-friendly* web-based tool with minimal technical requirements. The tool only requires a good internet connection, good quality MRI scans in the appropriate format and key information about the images from the user. This makes the tool suitable for non-experts in imaging and/or normative modelling. Secondly, Neurofind requires no prior *processing* of the images from the user. The only preprocessing necessary is to convert the scan into the required format (i.e., NIfTI), which is already standard practice in the field and accessible to most researchers. Thirdly, Neurofind can be applied to *any adult population*. By using whole-brain data across a large age range, Neurofind can be used to investigate brain morphology of psychiatric diagnoses. This differs from existing tools that are targeted at neurological disorders and therefore focus on relevant metrics and/or narrower age range such as gray matter of sub-cortical structures to detect AD or white matter tracts to detect multiple sclerosis. Finally, the tool

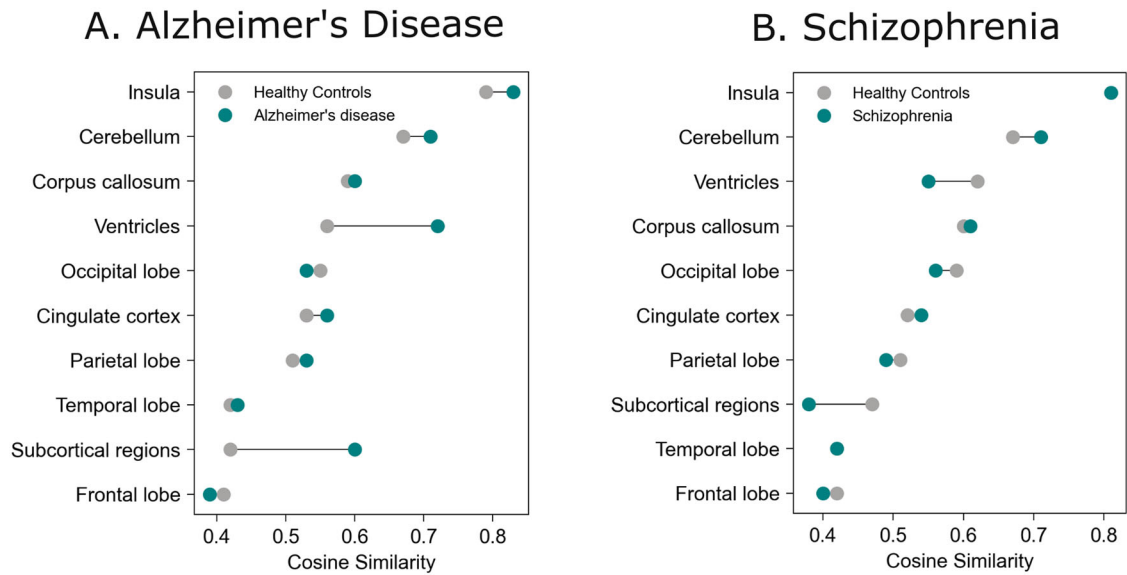


Fig. 7 Within-group mean pairwise cosine similarity for high-level brain regions. **A** Alzheimer's disease patients exhibited greater homogeneity in brain morphology compared to healthy controls, with the highest similarity observed in the ventricles and subcortical regions. **B** In contrast, Schizophrenia patients were slightly more heterogeneous than their respective controls, with the greatest differences in cosine similarity found in the ventricles and subcortical regions.

provides *two complementary normative metrics*. Interest in anomaly detection and brain age is growing rapidly in translational brain sciences. Making these models easily accessible to the wider research community will hopefully accelerate the field of normative modelling in psychiatric and neurological research. Notably, both metrics are transdiagnostic and derived from whole-brain data, allowing Neurofind to be applied to a group of individuals of interest without making any assumptions about their diagnoses. This will hopefully facilitate research on unique and shared brain morphology between diagnoses for example, as encouraged by dimensional approaches to psychopathology such as the Research Domain Criteria [79].

To illustrate the potential of Neurofind to investigate different aspects of any disorder of interest, we provided examples of possible analyses in AD and SZ. As expected, we found that AD patients have larger overall deviations from the normative brain morphology than comparable controls. Deviations were more pronounced in the ventricles and limbic-temporal regions. Alterations in these regions are well documented in literature on AD [81, 82]. In addition, the magnitude of the deviation from the norm for two of these regions - hippocampus and middle temporal gyrus - were negatively associated severity of dementia, such that patients with larger deviations from the norm presented with worse symptoms. We also showed that AD patients tended to be more homogeneous than comparable controls. This is an interesting finding that highlights the well-established morphological changes in AD. Consistent with the results from the group comparisons, homogeneity between patients was most pronounced in the ventricles and subcortical regions, which further emphasises the role of these brain regions in AD. In the analysis of brain age, AD patients had an older-appearing brain than comparable healthy controls consistent with the literature [46]. The results for SZ were not as clear cut. This is not unexpected given that, compared with AD, SZ has less clear morphological markers [83]. This may explain why the overall deviation from the norm was not statistically different from comparable controls, while specific regions namely the hippocampus and putamen were. These findings are also consistent with well documented brain alterations in SZ. Abnormal hippocampal volume in SZ is thought to be implicated in cognitive [84–86] and emotional processing [87, 88] deficits as well as in the dysregulation of the

hypothalamic-pituitary-adrenal axis [89]. Similarly, the pallidum also plays a central role in the dopaminergic dysfunction in SZ [90]. Finally, SZ patients tended to be more heterogeneous than controls across all regions and most of the high-level regions. Inter-individual variation in SZ is well-known but only recently have there been systematic efforts to quantify it [2, 91–93]. SZ patients also showed signs of accelerated ageing consistent with previous studies [46]. Taken collectively, these results indicate that Neurofind is capable of identifying meaningful deviations from the norm and can be a useful tool to parse heterogeneity in brain-based disorders. The less clear-cut results for schizophrenia are consistent with the literature on normative modelling, showing that only a small portion of patients fall outside the norm and that patients are highly heterogeneous in how they deviate from the expected [16, 18, 94]. Our aim with these exemplar analyses was to illustrate how Neurofind can be used to explore psychiatric and neurological disorders. There are many ways in which Neurofind can be explored further. For example, clustering patients based on region-level deviations, identifying differences in clinical presentation between within the norm and extreme patients, stratifying patients in terms of longitudinal outcomes, or investigate the association of brain deviations and behavioural/cognitive/health traits in healthy controls, such as BMI, substance use, cognitive performance. We encourage researchers to always use Neurofind in patients in combination with a group of comparable healthy controls, as shown in our examples, to control for biases specific to their sample. Inferences in patients scores should then be made in relation this comparison group.

The development of a practical tool for the wider research community, such as Neurofind, comes with a number of limitations and assumptions. Neurofind was designed to make normative modelling accessible and therefore it is not customisable. Advanced users may prefer to have more control over certain aspects of the data and/or model training. In addition, Neurofind was developed to provide general purpose and transdiagnostic metrics that can be used across psychiatric and neurological research. It was not developed to identify specific disorders or predict clinical outcomes in a real-world clinical setting. To maintain simplicity and interpretability of the findings, the brain age model in Neurofind did not take its statistical dependency on chronological age into account. Because of this

so-called 'regression to the mean' or 'age bias', the ages of younger subjects tend to be overestimated, while the ages of older subjects tend to be underestimated. We recommend that any further analysis of brain age uses age as a covariate or addresses this issue in another way [95]. Neuroharmony, the method used to minimise the impact of scanner-related bias, relies on a learned association between image quality metrics and needed corrections. For Neuroharmony to work effectively, it is necessary that the quality metrics of the images fall within a certain range; conversely, Neuroharmony may not be able to predict the required adjustments if the quality metrics are unusual (for more information see [56]). Similarly, Neurofind may not be suitable for analysis of subjects that are very different from our training range, i.e., outside of ages 20–80, non-white ethnicities, or MRI scans acquired with fields of strength considerably lower or higher than 1.5 T or 3 T. Finally, it is important to note that Neurofind is a research tool, so results should not be used to make clinical decisions.

In conclusion, Neurofind is a new freely available tool that aims to facilitate research in normative modelling in psychiatry and neurology. It is aimed at the wider community and non-expert researchers who wish to use this approach in their research. It relies on sound methods previously published on data harmonisation, brain age algorithms and deep learning. We have presented exemplar applications showing that Neurofind can produce interpretable and meaningful results in line with the literature on AD and SZ. We have also provided a how-to guide that illustrates the use Neurofind in four simple steps. Neurofind can be accessed via www.neurofind.ai and relevant publications and materials that describe the scanner harmonization, AAE and brain age methods in detail are also available [56, 59, 60].

DATA AVAILABILITY

All data used in this study were obtained from pre-existing datasets with varying access procedures. Researchers can refer to the individual dataset links below for further details: AIBL (<https://aibl.csiro.au/adni/index.html>), COBRE (https://fcon_1000.projects.nitrc.org/indi/retro/cobre.html), Human Connectome Project Aging (<https://www.humanconnectome.org/study/hcp-lifespan-aging/data-releases>), Human Connectome Project Young Adults (<https://www.humanconnectome.org/study/hcp-young-adult/data-releases>), IXI (<http://brain-development.org/ixi-dataset/>), MCIC (www.mmc.org), UK Biobank (<https://www.ukbiobank.ac.uk/enable-your-research>).

CODE AVAILABILITY

The code for model development is available at https://github.com/MLMH-Lab/neurofind_ai.

REFERENCES

- Whiteford HA, Ferrari AJ, Degenhardt L, Feigin V, Vos T. The global burden of mental, neurological and substance use disorders: an analysis from the global burden of disease study 2010. *PLoS One*. 2015;10:e0116820.
- Baldwin H, Radua J, Antoniades M, Haas SS, Frangou S, Agartz I, et al. Neuroanatomical heterogeneity and homogeneity in individuals at clinical high risk for psychosis. *Transl Psychiatry*. 2022;12:297.
- Kandilarova S, Stoyanov D, Sirakov N, Maes M, Specht K. Reduced grey matter volume in frontal and temporal areas in depression: contributions from voxel-based morphometry study. *Acta Neuropsychiatr*. 2019;31:252–7.
- Vieira S, Gong Q, Scarpazza C, Lui S, Huang X, Crespo-Facorro B, et al. Neuroanatomical abnormalities in first-episode psychosis across independent samples: a multi-centre mega-analysis. *Psychol Med*. 2021;51:340–50. <https://doi.org/10.1017/S0033291719003568>.
- Li T, van Rooij D, Roth Mota N, Buitelaar JK, Group TEAW, Hoogman M, et al. Characterizing neuroanatomic heterogeneity in people with and without ADHD based on subcortical brain volumes. *J Child Psychol Psychiatry*. 2021;62:1140–9.
- Shan X, Uddin LQ, Xiao J, He C, Ling Z, Li L, et al. Mapping the heterogeneous brain structural phenotype of autism spectrum disorder using the normative model. *Biol Psychiatry*. 2022;91:967–76.
- Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry*. 2012;17:1174–9.

- Sui J, Jiang R, Bustillo J, Calhoun V. Neuroimaging-based individualized prediction of cognition and behavior for mental disorders and health: methods and promises. *Biol Psychiatry*. 2020;88:818–28.
- Iniesta R, Stahl D, McGuffin P. Machine learning, statistical learning and the future of biological research in psychiatry. *Psychol Med*. 2016;46:2455–65.
- Bzdok D, Meyer-Lindenberg A. Machine learning for precision psychiatry: opportunities and challenges. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2018;3:223–30.
- Woo C-W, Chang LJ, Lindquist MA, Wager TD. Building better biomarkers: brain models in translational neuroimaging. *Nat Neurosci*. 2017;20:365–77.
- Marquand AF, Kia SM, Zabihi M, Wolfers T, Buitelaar JK, Beckmann CF. Conceptualizing mental disorders as deviations from normative functioning. *Mol Psychiatry*. 2019;24:1415–24.
- Verdi S, Marquand AF, Schott JM, Cole JH. Beyond the average patient: How neuroimaging models can address heterogeneity in dementia. *Brain*. 2021;144:2946–53.
- Marquand AF, Wolfers T, Mennes M, Buitelaar J, Beckmann CF. Beyond lumping and splitting: a review of computational approaches for stratifying psychiatric disorders. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2016;1:433–47. <https://doi.org/10.1016/j.bpsc.2016.04.002>.
- Marquand AF, Rezek I, Buitelaar J, Beckmann CF. Understanding heterogeneity in clinical cohorts using normative models: beyond case-control studies. *Biol Psychiatry*. 2016;80:552–61.
- Li J, Di Biase M, Cash RFH, Cocchi L, Cropley VL, Klauser P, et al. Individual deviations from normative models of brain structure in a large cross-sectional schizophrenia cohort. *Mol Psychiatry*. 2021;26:3512–23.
- Antoniades M, Haas SS, Modabbernia A, Bykowsky O, Frangou S, Borgwardt S, et al. Personalized estimates of brain structural variability in individuals with early psychosis. *Schizophr Bull*. 2021;47:1029–38. <https://doi.org/10.1093/schbul/sbab005>.
- Wolfers T, Doan NT, Kaufmann T, Alnaes D, Moberget T, Agartz I, et al. Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. *JAMA Psychiatry*. 2018;75:1146.
- Wolfers T, Beckmann CF, Hoogman M, Buitelaar JK, Franke B, Marquand AF. Individual differences v. the average patient: mapping the heterogeneity in ADHD using normative models. *Psychol Med*. 2019;50:314–23.
- Wolfers T, Rokicki J, Alnaes D, Agartz I, Kia SM, Kaufmann T et al. Extensive brain structural heterogeneity in individuals with schizophrenia and bipolar disorder. *medRxiv* 2020. <https://doi.org/10.1101/2020.05.08.20095091>.
- Liu Z, Palaniyappan L, Wu X, Zhang K, Du J, Zhao Q, et al. Resolving heterogeneity in schizophrenia through a novel systems approach to brain structure: individualized structural covariance network analysis. *Mol Psychiatry*. 2021;26:7719–31.
- Maron-Katz A, Zhang Y, Narayan M, Wu W, Toll RT, Naparstek S, et al. Individual patterns of abnormality in resting-state functional connectivity reveal two data-driven PTSD subgroups. *Am J Psychiatry*. 2020;177:244–53.
- Scarpazza C, Ha M, Baecker L, Garcia-Dias R, Pinaya WHL, Vieira S, et al. Translating research findings into clinical practice: a systematic and critical review of neuroimaging-based clinical tools for brain disorders. *Transl Psychiatry*. 2020;10:107. <https://doi.org/10.1038/s41398-020-0798-6>.
- Morra JH, Tu Z, Apostolova LG, Green AE, Avedissian C, Madsen SK, et al. Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls. *Neuroimage*. 2008;43:59–68.
- Suppa P, Anker U, Spies L, Bopp I, Ruegger-Frey B, Klaghofer R, et al. Fully automated atlas-based hippocampal volumetry for detection of Alzheimer's disease in a memory clinic setting. *J Alzheimers Dis*. 2015;44:183–93.
- Suppa P, Hampel H, Spies L, Fiebach JB, Dubois B, Buchert R, et al. Fully automated atlas-based hippocampus volumetry for clinical routine: validation in subjects with mild cognitive impairment from the ADNI cohort. *J Alzheimers Dis*. 2015;46:199–209.
- Brewer JB, Magda S, Airriess C, Smith ME. Fully-automated quantification of regional brain volumes for improved detection of focal atrophy in Alzheimer disease. *AJNR Am J Neuroradiol*. 2009;30:578–80.
- Vrooman HA, Cocosco CA, van der Lijn F, Stokking R, Ikram MA, Vernooij MW, et al. Multi-spectral brain tissue segmentation using automatically trained k-nearest-neighbor classification. *Neuroimage*. 2007;37:71–81.
- Manjón JV, Coupé P. volBrain: an online MRI brain volumetry system. *Front Neuroinform*. 2016;10:30. <https://doi.org/10.3389/fninf.2016.00030>.
- Romero JE, Coupé P, Manjón JV. HIPS: a new hippocampus subfield segmentation method. *Neuroimage*. 2017;163:286–95.
- Spies L, Tewes A, Suppa P, Opfer R, Buchert R, Winkler G, et al. Fully automatic detection of deep white matter T1 hypointense lesions in multiple sclerosis. *Phys Med Biol*. 2013;58:8323–37.
- de Boer R, Vrooman HA, van der Lijn F, Vernooij MW, Ikram MA, van der Lugt A, et al. White matter lesion extension to automatic brain tissue segmentation on MRI. *Neuroimage*. 2009;45:1151–61.

33. Coupé P, Tourdias T, Linck P, Romero JE, Manjon JV. LesionBrain: An online tool for white matter lesion segmentation. *Lect. Notes Comput. Sci.*, Springer 95–103 (2018).
34. Jain S, Vyver TV, Terzopoulos V, Sima DM, Roura E, Maas A, et al. Automatic quantification of computed tomography features in acute traumatic brain injury. *J Neurotrauma*. 2019;36:1794–803.
35. Brezova V, Moen KG, Skandsen T, Vik A, Brewer JB, Salvesen O, et al. Prospective longitudinal MRI study of brain volumes and diffusion changes during the first year after moderate to severe traumatic brain injury. *Neuroimage Clin*. 2014;5:128–40.
36. Ochs AL, Ross DE, Zannoni MD, Abildskov TJ, Bigler ED, Alzheimer's Disease Neuroimaging Initiative. Comparison of automated brain volume measures obtained with NeuroQuant and FreeSurfer. *J Neuroimaging*. 2015;25:721–7.
37. Ross DE, Ochs AL, Seabaugh JM, Shrader CR, Alzheimer's Disease Neuroimaging Initiative. Man versus machine: comparison of radiologists' interpretations and NeuroQuant® volumetric analyses of brain MRIs in patients with traumatic brain injury. *J Neuropsychiatry Clin Neurosci*. 2013;25:32–39.
38. Ross DE, Ochs AL, DeSmit ME, Seabaugh JM, Havranek MD, Alzheimer's Disease Neuroimaging Initiative. Man versus machine part 2: comparison of radiologists' interpretations and NeuroQuant measures of brain asymmetry and progressive atrophy in patients with traumatic brain injury. *J Neuropsychiatry Clin Neurosci*. 2015;27:147–52.
39. Azab M, Carone M, Ying SH, Yousem DM. Mesial temporal sclerosis: accuracy of NeuroQuant versus neuroradiologist. *AJNR Am J Neuroradiol*. 2015;36:1400–6.
40. Farid N, Girard HM, Kemmotsu N, Smith ME, Magda SW, Lim WY, et al. Temporal lobe epilepsy: quantitative MR volumetry in detection of hippocampal atrophy. *Radiology*. 2012;264:542–50.
41. Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, et al. Deep learning algorithms for detection of critical findings in head CT scans: A retrospective study. *Lancet*. 2018;392:2388–96.
42. Vieira S, Pinaya WHL, Mechelli A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications. *Neurosci Biobehav Rev*. 2017;74:58–75.
43. Pinaya WHL, Mechelli A, Sato JR. Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: A large-scale multi-sample study. 2018. <https://doi.org/10.1002/hbm.24423>.
44. Lopez Pinaya WH, Vieira S, Garcia-Dias R, Mechelli A. Autoencoders. 2019. <https://doi.org/10.1016/B978-0-12-815739-8.00011-0>.
45. Cole JH, Franke K. Predicting age using neuroimaging: innovative brain ageing biomarkers. *Trends Neurosci*. 2017;40:681–90.
46. Baecker L, Garcia-Dias R, Vieira S, Scarpazza C, Mechelli A. Machine learning for brain age prediction: introduction to methods and clinical applications. *EBio-Medicine*. 2021;72:1–9.
47. Chung Y, Addington J, Bearden CE, Cadenhead K, Cornblatt B, Mathalon DH, et al. Use of machine learning to determine deviance in neuroanatomical maturity associated with future psychosis in youths at clinically high risk. *JAMA Psychiatry*. 2018;75:960–8.
48. Koutsouleris N, Davatzikos C, Borgwardt S, Gaser C, Bottlender R, Frodl T, et al. Accelerated brain aging in schizophrenia and beyond: a neuroanatomical marker of psychiatric disorders. *Schizophr Bull*. 2013;40:1140–53.
49. Schnack HG, van Haren NEM, Nieuwenhuis M, Hulshoff Pol HE, Cahn W, Kahn RS. Accelerated brain aging in schizophrenia: A longitudinal pattern recognition study. *Am J Psychiatry*. 2016;173:607–16.
50. Hajek T, Franke K, Kolenic M, Capkova J, Matejka M, Propper L, et al. Brain age in early stages of bipolar disorders or schizophrenia. *Schizophr Bull*. 2017;45:190–8.
51. Kaufmann T, van der Meer D, Doan NT, Schwarz E, Lund MJ, Agartz I, et al. Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nat Neurosci*. 2019;22:1617–23.
52. Shahab S, Mulsant BH, Levesque ML, Calarco N, Nazeri A, Wheeler AL, et al. Brain structure, cognition, and brain age in schizophrenia, bipolar disorder, and healthy controls. *Neuropsychopharmacology*. 2019;44:898–906.
53. Van Gestel H, Franke K, Petite J, Slaney C, Garnham J, Helmick C, et al. Brain age in bipolar disorders: effects of lithium treatment. *Aust N Z J Psychiatry*. 2019;53:1179–88.
54. Beheshti I, Maikusa N, Matsuda H. The association between “Brain-Age Score” (BAS) and traditional neuropsychological screening tools in Alzheimer's disease. *Brain Behav*. 2018;8:1–14.
55. Varikuti DP, Genon S, Sotiras A, Schwender H, Hoffstaedter F, Patil KR, et al. Evaluation of non-negative matrix factorization of grey matter in age prediction. *Neuroimage*. 2018;173:394–410.
56. Garcia-Dias R, Scarpazza C, Baecker L, Vieira S, Pinaya WHL, Corvin A, et al. Neuroharmony: a new tool for harmonizing volumetric MRI data from unseen scanners. *Neuroimage*. 2020;220:117127.
57. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8:118–27.
58. Fortin J-P, Cullen N, Sheline YI, Taylor WD, Aselcioglu I, Cook PA, et al. Harmonization of cortical thickness measurements across scanners and sites. *Neuroimage*. 2018;167:104–20.
59. Pinaya WHL, Scarpazza C, Garcia-Dias R, Vieira S, Baecker L, F da Costa P, et al. Using normative modelling to detect disease progression in mild cognitive impairment and Alzheimer's disease in a cross-sectional multi-cohort study. *Sci Rep*. 2021;11:15746.
60. Baecker L, Dafflon J, da Costa PF, Garcia-Dias R, Vieira S, Scarpazza C, et al. Brain age prediction: a comparison between machine learning models using region- and voxel-based morphometric data. *Hum Brain Mapp*. 2021;42:2332–46.
61. Van Essen DC, Smith SM, Barch DM, Behrens TEJ, Yacoub E, Ugurbil K. The WU-minn human connectome project: an overview. *Neuroimage*. 2013;80:62–79.
62. Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, et al. The minimal preprocessing pipelines for the human connectome project. *Neuroimage*. 2013;80:105–24.
63. Bookheimer SY, Salat DH, Terpstra M, Ances BM, Barch DM, Buckner RL, et al. The lifespan human connectome project in aging: an overview. *Neuroimage*. 2019;185:335–48.
64. Miller KL, Alfaro-almagro F, Bangerter NK, Thomas DL, Yacoub E, Xu J, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci*. 2016;19:1523–36. <https://doi.org/10.1038/nn.4393>.
65. Esteban O, Birman D, Schaer M, Koyejo OO, Poldrack RA, Gorgolewski KJ. MRIQC: advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS One*. 2017;12:1–21.
66. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*. 2002;33:341–55.
67. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*. 2006;31:968–80.
68. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2006;8:118–27.
69. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825–30.
70. Ellis KA, Bush AI, Darby D, De Fazio D, Foster J, Hudson P, 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*. 2009;21:672–87.
71. Ellis KA, Rowe CC, Villemagne VL, Martins RN, Masters CL, Salvado O, et al. Addressing population aging and Alzheimer's disease through the Australian Imaging Biomarkers and Lifestyle study: collaboration with the Alzheimer's disease neuroimaging initiative. *Alzheimers Dement*. 2010;6:291–6.
72. Gollub RL, Shoemaker JM, King MD, White T, Ehrlich S, Sponheim SR, et al. The MCIC collection: a shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia. *Neuroinformatics*. 2013;11:367–88.
73. Mayer AR, Ruhl D, Merideth F, Ling J, Hanlon FM, Bustillo J, et al. Functional imaging of the hemodynamic sensory gating response in schizophrenia. *Hum Brain Mapp*. 2013;34:2302–12.
74. Tombaugh TN, McIntyre NJ. The mini-mental state examination: a comprehensive review. *J Am Geriatr Soc*. 1992;40:922–35.
75. Andreasen NC. The scale for the assessment of negative symptoms (SANS): conceptual and theoretical foundations. *Br J Psychiatry*. 1989;155:49–52.
76. Andreasen NC. Scale for the assessment of positive symptoms (SAPS). University of Iowa Press; Iowa City: 1984.
77. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull*. 1987;13:261–76.
78. Rutherford S, Barkema P, Tso IF, Sripada C, Beckmann CF, Ruhe HG, et al. Evidence for embracing normative modeling. *eLife*. 2023;12:e85082.
79. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry*. 2010;167:748–51.
80. Mechelli A, Vieira S. From models to tools: clinical translation of machine learning studies in psychosis. *NPJ Schizophr*. 2020;6:4.
81. Frisoni GB, Fox NC, Jack CR, Scheltens P, Thompson PM. The clinical use of structural MRI in Alzheimer disease. *Nat Rev Neurol*. 2010;6:67–77.
82. Bartos A, Gregus D, Ibrahim I, Tintêra J. Brain volumes and their ratios in Alzheimer's disease on magnetic resonance imaging segmented using FreeSurfer 6.0. *Psychiatry Res Neuroimaging*. 2019;287:70–74.
83. Prata D, Mechelli A, Kapur S. Clinically meaningful biomarkers for psychosis: a systematic and quantitative review. *Neurosci Biobehav Rev*. 2014;45:134–41.
84. Kuperberg G, Heckers S. Schizophrenia and cognitive function. *Curr Opin Neurobiol*. 2000;10:205–10.
85. Buchy L, Czechowska Y, Chochol C, Malla A, Joobar R, Pruessner J, et al. Toward a model of cognitive insight in first-episode psychosis: verbal memory and hippocampal structure. *Schizophr Bull*. 2010;36:1040–9.

86. Olypher AV, Klement D, Fenton AA. Cognitive disorganization in hippocampus: A physiological model of the disorganization in psychosis. *J Neurosci*. 2006;26:158–68.
87. Gur RE, McGrath C, Chan RM, Schroeder L, Turner T, Turetsky BI, et al. An fMRI study of facial emotion processing in patients with schizophrenia. *Am J Psychiatry*. 2002;159:1992–9.
88. Kohler CG, Martin EA. Emotional processing in schizophrenia. *Cognit Neuropsychiatry*. 2006;11:250–71.
89. Pruessner M, Cullen AE, Aas M, Walker EF. The neural diathesis-stress model of schizophrenia revisited: an update on recent findings considering illness stage and neurobiological and methodological complexities. *Neurosci Biobehav Rev*. 2017;73:191–218.
90. McCutcheon RA, Abi-Dargham A, Howes OD. Schizophrenia, dopamine and the striatum: from biology to symptoms. *Trends Neurosci*. 2019;42:205–20.
91. Voineskos AN, Jacobs GR, Ameis SH. Neuroimaging heterogeneity in psychosis: neurobiological underpinnings and opportunities for prognostic and therapeutic innovation. *Biol Psychiatry*. 2020;88:95–102.
92. Brugger SP, Howes OD. Heterogeneity and homogeneity of regional brain structure in schizophrenia. *JAMA Psychiatry*. 2017;74:1104.
93. Alnaes D, Kaufmann T, van der Meer D, Córdova-Palomera A, Rokicki J, Moberget T, et al. Brain heterogeneity in schizophrenia and its association with polygenic risk. *JAMA Psychiatry*. 2019;76:739–48.
94. Elad D, Cetin-Karayumak S, Zhang F, Cho KIK, Lyall AE, Seitz-Holland J, et al. Improving the predictive potential of diffusion MRI in schizophrenia using normative models—towards subject-level classification. *Hum Brain Mapp*. 2021;42:4658–70.
95. de Lange A-MG, Anatórk M, Rokicki J, Han LKM, Franke K, Alnaes D, et al. Mind the gap: performance metric evaluation in brain-age prediction. *Hum Brain Mapp*. 2022;43:3113–29.

ACKNOWLEDGEMENTS

This project was funded by a grant from Wellcome's Psychosis Flagship Innovations (220402/Z/20/Z) awarded to AM. SV was supported by a Sir Henry Wellcome postdoctoral fellowship (221638/Z/20/Z). This research has been conducted using the UK Biobank Resource (40323), Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657; funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research, and by the McDonnell Center for Systems Neuroscience at Washington University), Human Connectome Project - Aging (National Institute On Aging of the National Institutes of Health; U01AG052564; <https://doi.org/10.15154/rad1-0h66>), IXI (<https://brain-development.org/ixi-dataset>), AIBL (<https://aibl.csiro.au>), COBRE (National Institute of Health Center of Biomedical Research Excellence; 1P20RR021938-01A2) and MCIC (<http://coins.mrn.org/dataexchange>). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

AUTHOR CONTRIBUTIONS

Based on the Contributor Roles Taxonomy (CRediT) authorship contribution guidance, we declare the following author contributions. SV: Writing - Original Draft, Visualization, Software, Formal analysis; LB: Writing - Review & Editing, Data Curation,

Visualization, Software, Formal analysis; WHLP: Writing - Review & Editing, Data Curation, Software, Methodology, Formal analysis; RGD: Writing - Review & Editing, Data Curation, Software, Methodology, Formal analysis; CS: Writing - Review & Editing, Data Curation, Conceptualization; VC: Writing - Review & Editing, Funding acquisition; AM: Writing - Review & Editing, Conceptualization, Funding acquisition.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS

The authors hereby confirm that all methodologies employed in this study adhere to the ethical guidelines set forth by the relevant national and institutional committees on human research, as well as align with the principles of the Helsinki Declaration. This study involved analysis of pre-existing anonymised datasets available elsewhere. The collection of all data was approved by the local ethics committees. Informed consent was obtained from all participants by the local research teams. Further details on data collection are provided in the references for the individual datasets in the methods. The analysis of the pre-existing subject data conducted as part of this study received full approval by the Psychiatry, Nursing and Midwifery Research Ethics Subcommittee at King's College London (reference number LRS-20/21-21260).

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41398-025-03290-x>.

Correspondence and requests for materials should be addressed to A. Mechelli.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025