Update

Beyond the average patient: how neuroimaging models can address heterogeneity in dementia

Serena Verdi,[1,2] Andre F. Marquand,[3,4] Jonathan M. Schott[2] and James H. Cole[1,2]

Abstract

Dementia is a highly heterogeneous condition, with pronounced individual differences in onset age, clinical presentation, progression rates and neuropathological hallmarks, even within a specific diagnostic group. However, the most common statistical designs used in dementia research studies and clinical trials overlook this heterogeneity, instead relying on the comparison of group average differences (e.g., patient versus control, treatment versus placebo), implicitly assuming within-group homogeneity. This one-size-fits-all approach potentially limits our understanding of dementia aetiology, hindering the identification of effective treatments. Neuroimaging has enabled characterisation of the average neuroanatomical substrates of dementias; however, the increasing availability of large open neuroimaging datasets provides the opportunity to examine patterns of neuroanatomical variability in individual patients. In this Update review we outline the causes and consequences of heterogeneity in dementia and discuss recent research which aims to directly tackle heterogeneity, rather than assume that dementia affects everyone in the same way. We introduce spatial normative modelling as an emerging data-driven technique which can be applied to dementia data to model neuroanatomical variation, capturing individualised neurobiological "fingerprints". Such methods have the potential to detect clinically relevant subtypes, track an individual's disease progression or evaluate treatment responses, with the goal of moving towards precision medicine for dementia.

Author affiliations:

1 Centre for Medical Image Computing, Medical Physics and Biomedical Engineering, University College London, London, UK

2 Dementia Research Centre, UCL Queen Square Institute of Neurology, London, UK

3 Donders Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behaviour, Radbound University, Nijmegen, 6525EN, The Netherlands

4 Department of Cognitive Neuroscience, Radbound University Medical Centre, Nijmegen, 6525EN, The Netherlands

Correspondence to: Serena Verdi

The Centre for Medical Image Computing

90 High Holborn, Floor 1, London, WC1V 6LJ, UK

E-mail: serena.verdi.19@ucl.ac.uk

Introduction

Heterogeneity is an underlying characteristic of dementia, in presentation and progression. Variability is observed in underlying neuropathology, in genetic risk factors, in imaging and fluid biomarkers and in clinical and behaviour manifestations; reinforcing the idea that each dementia patient is unique. However, it is challenging to capture this heterogeneity when studying dementia and it is often not attempted. The conventional analytical approach focuses on characterising group averages, not individual differences, assuming homogeneity between dementia patients **(Fig.1, Panel A)**. The failure to incorporate heterogeneity in statistical models of dementia may have limited our understanding of the pathophysiological mechanisms and slowed the development of treatments.

Despite thousands of treatment trials, only one drug (aducanumab) shows any promise for disease modification.[1] This paucity of treatments, in combination with the rapid ageing of the global population, adds to the societal burden of dementia.[2] This motivates re-evaluation of common experimental approaches in dementia research and clinical trials with the goal of optimising statistical design. In this Update, we review current and emerging neuroimaging analysis methods which can account for the intrinsic heterogeneity to help further our understanding of dementia and improve the prospect of developing effective treatments.

Heterogeneity in dementia

Dementia is characterised by progressive cognitive decline, over and above that seen in normal ageing, with subsequent impact on activities of daily living. Dementia is the endpoint of

multiple diseases, including Alzheimer's Disease (AD), vascular dementia (VD), Parkinson's Disease Dementia (PDD), Dementia with Lewy Bodies (DLB), Frontotemporal Dementia (FTD), and Limbic-predominant Age-related TDP-43 Encephalopathy (LATE).[3,4] Neuropathological factors vary between these diseases, though commonly include amyloid, tau and Lewy body accumulation[5,6] and neurodegeneration.

The neuropathological hallmarks of the dementia syndromes have been well characterised in some cases, while others continue to be defined. Indeed, the Braak stages are an established method for characterising and defining AD and PDD pathology.[7,8] However, neurobiological evidence suggests that there are many individual exceptions to these general rules. This could be because of the complex relationship between clinical syndrome and underlying neuropathology, as well as individual differences in brain structure and function that predate pathological onset. For instance, neocortical neuritic amyloid plaques have been observed post-mortem in over 50% of non-demented older adults, while almost 20% of dementia patients had no such plaques at death.[9] Beyond amyloid, other dementia risk factors such as tau-tangles, white-matter lesions and vascular pathologies have been recorded in dementia-free older adults.[9] This suggests that these pathological hallmarks of dementia are not universal, with individuals are resilient to insults that may be sufficient to cause dementia in others.

The mechanisms driving these pathological changes are yet to be fully determined, though likely vary between individuals, both between and within diagnostic categories. Genetic and environmental risk factors also show considerable variability here.[10,11] For example, APOE is the best-known genetic risk for sporadic AD, but is only semi-dominant and moderately penetrant; at age 85 years between 30-50% of APOE4 homozygotes do not have dementia.[12] Potentially, different genetic and molecular mechanisms (or combinations of mechanisms) can result in dementia.

In addition, separate pathologies have broad phenotypic correspondence and clinicopathological relationships can be varied (e.g., aphasia, behavioural disturbance in FTLD).[13] A specific dementia phenotype may be the result of different pathological processes, and conversely, a single molecular pathology may result in multiple different dementia phenotypes.[14] A specific diagnosis (e.g., AD) can also include pathological features which are characteristic of another dementia disease (e.g., TDP-43 proteinopathy).[15] Symptoms often do not conform to diagnostic boundaries.[16,17] For instance, diverse symptoms are observed in frontotemporal lobar degeneration syndromes, but do not easily fall within existing clinical

categories or a single disease entity.[18] Furthermore, heterogeneity can be seen in the severity of symptoms,[19] rates of change,[20] and influence on activities of daily life.[21]

To add further complexity, many dementia patients have non-dementia comorbidities, for example, neuropsychiatric and gastrointestinal diagnoses,[22] all of which may impact clinical presentation and disease progression. Age is also a key risk factor for pathophysiological changes; disentangling disease-related variation from the ageing process is challenging, for example when differentiating between normal cognitive decline or MCI as a prodromal phase of AD.[23]

Neuroimaging can offer *in vivo* neurobiological insights into the inter-individual variability in dementia.[24–26] Structural MRI has uncovered anatomical differences in dementia patients, reflected in pattens of atrophy,[27] and anatomical symmetry.[28] Differences in pathological hallmarks, can be seen using PET ligands, for instance tau (e.g., ^{18}F-AV1451) and amyloid (e.g., ^{11}C-PiB) ligand binding has shown to vary in within AD patients.[29] Functional MRI has also been used to capture differences in connectivity.[30]

Neuroimaging is now commonly implemented in clinical trials to provide secondary outcome measures of treatment effectiveness.[31–33] Here, the relationship between variance and statistical power is considered; variance can increase noise in the typical case-control designs and subsequently reduce statistical power to detect change. For example, in early trials of solanezumab, both amyloid-negative and amyloid-positive AD patients (defined using PET) were recruited, reducing the sensitivity to detect effects as some patients were likely following different pathological trajectories.[34,35] Subsequently, it has been argued that a stringent approach of selecting the optimal participants and the biomarker of interest for clinical trials will increase the likelihood of success when evaluating average differences between patients.[36] However, there are limitations to this approach. For example, in PDD hippocampal involvement is widely disputed,[37] and in AD both higher,[38] or lower,[39] caudate nucleus volumes have been reported compared to healthy controls. Such inconsistencies could be due to differences within patient groups in either disease subtype or disease stage. Until both disease heterogeneity and disease dynamics are better understood, using biomarkers for stratified trial enrolment will likely remain contentious.

A key limitation with current statistical approaches is the assumption in traditional case-control studies that experimental groups are homogenous discrete entities **(Fig.1, Panel A).** Here, tests of statistical significance are based on group means, generally regarding individual differences

as error or noise. In other words, this approach is fundamentally oriented to comparing the 'average patient'. Even in sophisticated multivariate analyses (e.g., machine learning), the focus tends to be on the discovery of canonical patterns across sets of variables that differentiate one group from another. This assumption of within-group homogeneity is neither reflected in real-world clinical populations nor in the heterogeneous pathological nature of neurodegenerative diseases. Inconsistencies are commonly seen in treatment effects,[2,40,41] however this could be related to unmeasured individual differences rather than poor efficacy. The assumption that there are uniform effects of dementia or of treatment on the brain may be hindering the discovery of disease-modifying treatments, especially when translating to heterogeneous clinical settings. This motivates the incorporation of heterogeneity in trial design.[42] Given the importance of the brain in dementia, we outline ways to model heterogeneity using neuroimaging and illustrate the impact this could have on fundamental research and clinical trials.

Data-driven statistical methods

Measuring and statistically modelling neurobiological heterogeneity in a clinical population requires large datasets. Fortunately, large neuroimaging datasets are increasingly available for dementia, such as Alzheimer's Disease Neuroimaging Initiative (ADNI), Open Access Series of Imaging Studies (OASIS) and National Alzheimer's Coordinating Center (NACC). These datasets allow for more flexible and powerful statistical testing.[43–45] These data have supported the development and application of novel data-driven methods designs in dementia research **(Box 1)**.

Recently, data-driven methods have enabled the estimation of disease subtypes from neuroimaging data, a promising way to disentangle heterogeneity by grouping patients by distinctive neurobiological and cognitive characteristics,[46] and disease progression.[47,48] For instance, hierarchical clustering algorithms have been utilised to understand variation in cortical thickness,[49,50] grey matter[51] and progressive neurodegeneration.[52] Clustering techniques employed on large AD neuroimaging datasets have suggested that there are disease subtypes with distinct patterns of cortical thinning. Atrophy-based groupings have been defined as either medial-temporal, parietal or widespread ('diffuse').[49,53] Interestingly, these subtypes have also been associated with patterns of both amyloid[50] and tau deposition [54] and with cognitive phenotypes.[55,56] In FTD, distinct atrophy subtypes have been reported, corresponding to the temporal-dominant, temporofrontoparietal, frontal-dominant and frontotemporal

areas.[51,57] In DLB, distinct atrophy subtypes have also been reported: non-atrophic, parietotemporal atrophy and occipitofrontal atrophy, with corresponding distinctions in cognitive performance.[58,59] Despite some consistency between different reports, conversely other studies have generated a different number of subtypes, and others that have reported subtypes with anatomical overlap, such as occipital areas overlapping with parietal and mild atrophic patterns.[60]

Most studies using clustering methods have analysed cross-sectional single time-point data, however, given the heterogeneity in disease progression, mapping longitudinal trajectories is an important focus for dementia research. Young and colleagues recently combined disease progression modelling and clustering techniques, to enable inference of subtype and disease stage.[52] Here three distinct spatiotemporal atrophy patterns were observed in AD, with atrophy starting in either the medial temporal lobe, frontotemporal areas or the basal ganglia. In addition, four distinct spatiotemporal atrophy patterns were observed in FTD, corresponding to different genetic subtypes.[52] Future efforts could continue to explore heterogeneity in disease progression, including examining presymptomatic and prodromal disease phases. Parsing this longitudinal heterogeneity should enable stratification of dementia patients into groups with differing disease progression rates, with treatments and interventions tailored to these groups accordingly.

Defining biologically meaningful subtypes may have implications for fundamental research. Case-control genome-wide association studies (GWAS) are hindered by heterogeneity in patients and controls alike,[61] and dementia is likely no except. Restricting GWAS to more homogeneous subtypes should increase the sensitivity of such research to reliability detect genetic risk factors for dementia.

Despite promising initial results from subtyping studies, there are key issues to consider prior to translating such models into clinical settings. The number of subtypes generated, subtype distinguishability and the stability of subtypes over the disease course should be considered.[46] It is possible that subtypes may be confounded by statistical decisions (e.g., hyperparameter choices), technical factors (e.g., scanner), biomedical factors (e.g., age, sex, comorbid disease) or sampling bias. Validation of clustering-derived subtypes is challenging in the absence of ground truth.[62] Therefore, to ensure subtypes are biologically meaningful, external validation steps using independent datasets and long-term clinical outcome data (i.e. mortality rates and post-mortem data)[63] are essential. Recently promising results have emerged using post-mortem

histopathological data, which have yielded transdiagnostic disease clusters.[64] However, the availability of datasets that enable such validation are limited, causing a bottleneck in the progress of dementia subtyping research. Hopefully, efforts to access existing hospital and community data will be successful in providing the data necessary for clinical validation.

It is also important to consider if discrete subtypes can explain the range of variability observed. By design, clustering assumes homogeneity within each cluster, which itself may not be valid.[65] This motivates research into less discrete, or overlapping subtypes, with multiple subtype factors contributing to explain patterns of brain structure of any particular individual, such as Zhang and colleagues' Bayesian latent factor analysis.[66]

Spatial normative modelling

Going further, it is possible to assess the neurobiology of dementia at the level of the individual patient and provide still greater precision than subtyping. To that end, normative modelling techniques have been developed to parse spatial heterogeneity (i.e., individual level regional variation) in neuroimaging data. Principally, normative modelling involves calculating the normal distribution of a population, then assessing how much an individual deviates from that respective distribution. Spatial normative modelling is a technique which specifically uses neuroimaging data (e.g., cortical thickness) to estimate variation for a given brain region.[67,68] This is detailed in **Box 2** and illustrated in **Fig. 2**. The extent to which an individual deviates from the norm can be spatially mapped at regions across the brain, providing an idiosyncratic map of individual variability. These 'z-score' maps can further be summarised to provide a patient-level index of deviation potentially reflecting their general brain health.

Multiple algorithms that have been proposed for spatial normative modelling. A common approach employs Gaussian processes regression to estimate normative models in the brain.[67] More recent developments have included a "neural processes" model, which does not rely on fixed parametric kernels and can improve the scaling of the model to large datasets. This method learns optimal feature representations and covariance structure for random-effect and noise (via global latent variables),[69] and a hierarchical Bayesian regression (HBR) approach to normative modelling has shown to efficiently accommodate inter-site variation and provides computational scaling, which is useful when using large studies, or combining smaller studies together, that are acquired across multiple sites.[70]

Preliminary studies have used normative modelling in the context of brain ageing and dementia. Distinct patterns of deviation from normal ranges were observed in people with MCI

and in AD patients.[71] Additionally, quantile regression techniques have been used to map deviations of cognitively normal individuals and AD patients brain morphology. Here differences between AD patients and healthy controls were partly attributed to accelerated aging.[72] Recently, a spatial pattern of atrophy index, reflecting normative variability, was used to demonstrate greater age-related atrophy in AD patients compared to normative trends of age-related changes in brain structure.[73]

Spatial normative modelling has also been applied to neuroimaging data in the contexts of attention-deficit hyperactivity disorder,[74] autism,[75,76] bipolar disorder and schizophrenia.[77] Results show that it is uncommon for patients to have uniform patterns of structural alterations across the brain. Individualised maps of regional differences derived from normative models generate distinct findings compared to case-control approaches, for example, using voxel-based morphometry,[75,76] which rely on modelling average differences between groups at each voxel. The predominant focus of neuroimaging research on group-level differences has potentially masked heterogeneity among patients within diagnostic groups.[74,77] Therefore, spatial normative modelling provides a new approach to examine the neurobiological correlates of neurodevelopmental and psychiatric disorders and could well be applied in dementia.

As spatial normative modelling can utilise any continuous or categorical phenotype, a range of dementia-related neuroimaging features (i.e., local volumes, cortical thickness, diffusion microstructural indices, PET tracer binding), could be used in the model to ascertain the heterogeneity across multiple aspects of neurobiology. In addition, a multimodal approach can be adopted (e.g., combining variables of neuroanatomical volumetric measures and PET amyloid beta binding) to parse heterogeneity at both a molecular and structural level.

Going beyond neuroimaging, other data can be used to disentangle the heterogeneity in dementia, such as fluid biomarkers, physiological measures, cognitive assessments, and genetic markers. Such information could be explicitly incorporated in spatial normative models as predictors (alongside age and sex) of brain structure. For instance, tau CSF levels are currently pivotal in diagnosis and treatment planning, therefore tau CSF could define a normative model of neuroanatomical measures to test whether elevated tau is associated with uniform or heterogeneous impacts on the brain. Alternatively, spatial normative models could be stratified, for example defining separate models for APOE4 carriers or people who are amyloid positive. While promising, these approaches require sufficiently large samples of normative data with that biomarker available, potentially challenging for invasive measures

like CSF sampling. Importantly, non-imaging markers provide a means of validating subtypes from neuroimaging normative models, which the assumption that neuroanatomically homogeneous subtypes would be more homogeneous in terms of genetic and environmental risks and fluid biomarker readouts.

Furthermore, spatial normative modelling could be particularly informative with longitudinal data. Temporal heterogeneity could be modelled using two or more time points to understand within-subject changes. In this instance, greater change in abnormality patterns could reflect a faster disease progression **(Fig.2, Panel D)**. Similarly, evidence of reduced or reversed changes in abnormality patterns could be indicative of treatment efficacy, using an outcome measure tailored to each individual's brain **(Fig.2, Panel E)**.

Summary

In conclusion, while dementia is associated with marked clinical, aetiological and neuropathological variability, research studies and clinical trials often overlook this inherent heterogeneity. While neuroimaging has provided many insights into the neuroanatomy of dementia and has helped assess treatment efficacy, the reliance on 'group-average' statistical methods may have hindered efforts to understand aetiology, to prognosticate for patients, and may account for failures in clinical drug development. We have outlined how data-driven neuroimaging statistical techniques enable explicit modelling of heterogeneity in the brain. We propose that the application of spatial normative modelling methods to dementia neuroimaging studies is a promising avenue to map regional variation at the individual level; efforts should include an investigation of various neuroanatomical markers derived from neuroimaging, which are then validated using independent datasets. Here, our clinicopathological understanding of anatomical variation could be enhanced by multimodal neuroimaging techniques and combining other biological data as predictors in normative models. Importantly, these spatial normative modelling techniques are not designed to replace or even improve on diagnosis based on clinical evidence and well-established biomarkers. The goal here instead is to better capture the variability *within* diagnostic groups based on individual patterns of brain structure, or potentially define neuroanatomical subtypes than span diagnostic boundaries. Employing these methods could be highly advantageous in mapping neurobiological abnormalities. Particularly, the use of serial neuroimaging to define patient level longitudinal trajectories of neuroanatomical variability has the potential to improve predictions of disease

progression or treatment response at the level of the individual patient and thereby paving the way towards more effective, precise medicine for dementia.

Competing interests

The authors report no competing interests

Funding

References

1.   Schneider L. A resurrection of aducanumab for Alzheimer's disease. *The Lancet Neurology*. 2020;19(2):111-112.

2.   Gauthier S, Albert M, Fox N, et al. Why has therapy development for dementia failed in the last two decades? *Alzheimer's and Dementia*. 2016;12(1):60-64.

3.   Burns A, Iliffe S. Alzheimer's disease. *BMJ (Online)*. 2009;338(7692):467-471.

4.   Nelson PT, Dickson DW, Trojanowski JQ, et al. Limbic-predominant age-related TDP-43 encephalopathy (LATE): Consensus working group report. *Brain*. 2019;142(6):1503-1527.

5.   Weller J, Budson A. Current understanding of Alzheimer's disease diagnosis and treatment. *F1000Research*. 2018;7.

6.   Armstrong MJ. Lewy body dementias. *CONTINUUM Lifelong Learning in Neurology*. 2019;25(1):128-146.

7.   Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica*. 1991;82(4):239-259.

8.   Braak H, Del Tredici K, Rüb U, De Vos RAI, Jansen Steur ENH, Braak E. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiology of Aging*. 2003;24(2):197-211.

9.   Savva GM, Wharton SB, Ince PG, Forster G, Matthews FE, Brayne C. Age, neuropathology, and dementia. *New England Journal of Medicine*. 2009;360(22):2302-2309.

10.   Cohen-Mansfield J. Heterogeneity in dementia: Challenges and opportunities. *Alzheimer Disease and Associated Disorders*. 2000;14(2):60-63.

11.   Ringman JM, Goate A, Masters CL, et al. Genetic Heterogeneity in Alzheimer Disease and Implications for Treatment Strategies. *Current Neurology and Neuroscience Reports*. 2014;14(11).

12.   Genin E, Hannequin D, Wallon D, et al. APOE and Alzheimer disease: A major gene with semi-dominant inheritance. *Molecular Psychiatry*. 2011;16(9):903-907.

13.   Lam B, Masellis M, Freedman M, Stuss DT, Black SE. Clinical, imaging, and pathological heterogeneity of the Alzheimer's disease syndrome. *Alzheimer's Research and Therapy*. 2013;5(1).

14.   Rabinovici GD, Carrillo MC, Forman M, et al. Multiple comorbid neuropathologies in the setting of Alzheimer's disease neuropathology and implications for drug development. *Alzheimer's and Dementia: Translational Research and Clinical Interventions*. 2017;3(1):83-91.

15.   Thomas DX, Bajaj S, McRae-McKee K, Hadjichrysanthou C, Anderson RM, Collinge J. Association of TDP-43 proteinopathy, cerebral amyloid angiopathy, and Lewy bodies with cognitive impairment in individuals with or without Alzheimer's disease neuropathology. *Scientific Reports*. 2020;10(1):1-9.

16.   Cunningham EL, McGuinness B, Herron B, Passmore AP. Dementia. *Ulster Medical Journal*. 2015;84(2):79-87.

17.   Piguet O. Neither white nor black: Embracing clinical variability in dementia diagnosis. *Brain*. Published online 2020.

18.   Murley AG, Coyle-Gilchrist I, Rouse MA, et al. Redefining the multidimensional clinical phenotypes of frontotemporal lobar degeneration syndromes. *Brain*. 2020;143(5):1555-1571.

19.   Ryan J, Fransquet P, Wrigglesworth J, Lacaze P. Phenotypic heterogeneity in dementia: A challenge for epidemiology and biomarker studies. *Frontiers in Public Health*. 2018;6.

20.   Goyal D, Tjandra D, Migrino RQ, Giordani B, Syed Z, Wiens J. Characterizing heterogeneity in the progression of Alzheimer's disease using longitudinal clinical and neuroimaging biomarkers. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*. 2018;10:629-637.

21.   Urwyler P, Stucki R, Rampa L, Müri R, Mosimann UP, Nef T. Cognitive impairment categorized in community-dwelling older adults with and without dementia using in-home sensors that recognise activities of daily living. *Scientific Reports*. 2017;7.

22.   Bunn F, Burn AM, Goodman C, et al. Comorbidity and dementia: A scoping review of the literature. *BMC Medicine*. 2014;12(1).

23.   Toepper M. Dissociating normal aging from Alzheimer's disease: A view from cognitive neuroscience. *Journal of Alzheimer's Disease*. 2017;57(2):331-352.

24.   Cash DM, Rohrer JD, Ryan NS, Ourselin S, Fox NC. Imaging endpoints for clinical trials in Alzheimer's disease. *Alzheimer's Research and Therapy*. 2014;6(9).

25.   Dong A, Toledo JB, Honnorat N, et al. Heterogeneity of neuroanatomical patterns in prodromal Alzheimer's disease: links to cognition, progression and biomarkers. *Brain*. 2017;140(3):735-747.

26.   Weiner MW, Veitch DP, Aisen PS, et al. Recent publications from the Alzheimer's Disease Neuroimaging Initiative: Reviewing progress toward improved AD clinical trials. *Alzheimer's and Dementia*. 2017;13(4):e1-e85.

27.   Whitwell JL, Dickson DW, Murray ME, et al. Neuroimaging correlates of pathologically defined subtypes of Alzheimer's disease: A case-control study. *The Lancet Neurology*. 2012;11(10):868-877.

28.   Mesulam MM, Weintraub S, Rogalski EJ, Wieneke C, Geula C, Bigio EH. Asymmetry and heterogeneity of Alzheimer's and frontotemporal pathology in primary progressive aphasia. *Brain*. 2014;137(4):1176-1192.

29.   Ossenkoppele R, Schonhaut DR, Schöll M, et al. Tau PET patterns mirror clinical and neuroanatomical variability in Alzheimer's disease. *Brain*. 2016;139(5):1551-1567.

30.   Khan W, Amad A, Giampietro V, et al. The heterogeneous functional architecture of the posteromedial cortex is associated with selective functional connectivity differences in Alzheimer's disease. *Human Brain Mapping*. 2020;41(6):1557-1572.

31.   Klunk WE, Engler H, Nordberg A, et al. Imaging Brain Amyloid in Alzheimer's Disease with Pittsburgh Compound-B. *Annals of Neurology*. 2004;55(3):306-319.

32.   Valenzuela M, Bartrés-Faz D, Beg F, et al. Neuroimaging as endpoints in clinical trials: Are we there yet? Perspective from the first Provence workshop. *Molecular Psychiatry*. 2011;16(11):1064-1066.

33.   Staffaroni AM, Ljubenkov PA, Kornak J, et al. Longitudinal multimodal imaging and clinical endpoints for frontotemporal dementia clinical trials. *Brain*. 2019;142(2):443-459.

34.   Siemers ER, Sundell KL, Carlson C, et al. Phase 3 solanezumab trials: Secondary outcomes in mild Alzheimer's disease patients. *Alzheimer's and Dementia*. 2016;12(2):110-120.

35.   Honig LS, Vellas B, Woodward M, et al. Trial of solanezumab for mild dementia due to Alzheimer's disease. *New England Journal of Medicine*. 2018;378(4):321-330.

36.   Cummings J, Feldman HH, Scheltens P. The 'rights' of precision drug development for

Alzheimer's disease. *Alzheimer's Research and Therapy*. 2019;11(1).

37. Weil RS, Hsu JK, Darby RR, Soussand L, Fox MD. Neuroimaging in Parkinson's disease dementia: connecting the dots. *Brain Communications*. 2019;1(1).

38. Persson K, Selbæk G, Brækhus A, Beyer M, Barca M, Engedal K. Fully automated structural MRI of the brain in clinical dementia workup. *Acta Radiologica*. 2017;58(6):740-747.

39. Jiji S, Smitha KA, Gupta AK, Pillai VPM, Jayasree RS. Segmentation and volumetric analysis of the caudate nucleus in Alzheimer's disease. *European Journal of Radiology*. 2013;82(9):1525-1530.

40. Cummings J, Aisen PS, Dubois B, et al. Drug development in Alzheimer's disease: The path to 2025. *Alzheimer's Research and Therapy*. 2016;8(1).

41. Cummings J, Lee G, Ritter A, Sabbagh M, Zhong K. Alzheimer's disease drug development pipeline: 2019. *Alzheimer's and Dementia: Translational Research and Clinical Interventions*. 2019;5:272-293.

42. Aisen PS, Cummings J, Jack CR, et al. On the path to 2025: Understanding the Alzheimer's disease continuum. *Alzheimer's Research and Therapy*. 2017;9(1).

43. Weiner MW, Aisen PS, Jack CR, et al. The Alzheimer's Disease Neuroimaging Initiative: Progress report and future plans. *Alzheimer's and Dementia*. 2010;6(3).

44. Marek K, Chowdhury S, Siderowf A, et al. The Parkinson's progression markers initiative (PPMI) – establishing a PD biomarker cohort. *Annals of Clinical and Translational Neurology*. 2018;5(12):1460-1477.

45. LaMontagne P, Benzinger T, Morris J, et al. OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease. *medRxiv*. Published online 2019.

46. Habes M, Grothe MJ, Tunc B, McMillan C, Wolk DA, Davatzikos C. Disentangling Heterogeneity in Alzheimer's Disease and Related Dementias Using Data-Driven Methods. *Biological Psychiatry*. 2020;88(1):70-82.

47. Oxtoby NP, Young AL, Cash DM, et al. Data-driven models of dominantly-inherited Alzheimer's disease progression. *Brain*. 2018;141(5):1529-1544.

48. Oxtoby NP, Leyland L, Aksman LM, et al. Sequence of clinical and neurodegeneration

events in Parkinson's disease progression. *Brain*. Published online 2021:1-14.

49.   Noh Y, Jeon S, Lee JM, et al. Anatomical heterogeneity of Alzheimer disease Based on cortical thickness on MRIs. *Neurology*. 2014;83(21):1936-1944.

50.   Hwang J, Kim CM, Jeon S, et al. Prediction of Alzheimer's disease pathophysiology based on cortical thickness patterns. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*. 2016;2:58-67.

51.   Whitwell JL, Przybelski SA, Weigand SD, et al. Distinct anatomical subtypes of the behavioural variant of frontotemporal dementia: A cluster analysis study. *Brain*. 2009;132(11):2932-2946.

52.   Young AL, Marinescu R V., Oxtoby NP, et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. *Nature Communications*. 2018;9(1).

53.   Na HK, Kang DR, Kim S, et al. Malignant progression in parietal-dominant atrophy subtype of Alzheimer's disease occurs independent of onset age. *Neurobiology of Aging*. 2016;47:149-156.

54.   Jeon S, Kang JM, Seo S, et al. Topographical heterogeneity of Alzheimer's disease based on MR imaging, tau PET, and amyloid PET. *Frontiers in Aging Neuroscience*. 2019;10(JUL).

55.   Park JY, Na HK, Kim S, et al. Robust Identification of Alzheimer's Disease subtypes based on cortical atrophy patterns. *Scientific Reports*. 2017;7.

56.   Sun N, Mormino EC, Chen J, Sabuncu MR, Yeo BTT. Multi-modal latent factor exploration of atrophy, cognitive and tau heterogeneity in Alzheimer's disease. *NeuroImage*. 2019;201.

57.   Whitwell JL, Xu J, Mandrekar J, et al. Frontal asymmetry in behavioral variant frontotemporal dementia: Clinicoimaging and pathogenetic correlates. *Neurobiology of Aging*. 2013;34(2):636-639.

58.   Uribe C, Segura B, Baggio HC, et al. Patterns of cortical thinning in nondemented Parkinson's disease patients. *Movement Disorders*. 2016;31(5):699-708.

59.   Uribe C, Segura B, Baggio HC, et al. Progression of Parkinson's disease patients' subtypes based on cortical thinning: 4-year follow-up. *Parkinsonism and Related Disorders*.

2019;64:286-292.

60.   Ten Kate M, Dicks E, Visser PJ, et al. Atrophy subtypes in prodromal Alzheimer's disease are associated with cognitive decline. *Brain*. 2018;141(12):3443-3456.

61.   Manchia M, Cullis J, Turecki G, Rouleau GA, Uher R, Alda M. The Impact of Phenotypic and Genetic Heterogeneity on Results of Genome Wide Association Studies of Complex Diseases. *PLoS ONE*. Published online 2013.

62.   Badhwar AP, Peggy McFall G, Sapkota S, et al. A multiomics approach to heterogeneity in Alzheimer's disease: Focused review and roadmap. *Brain*. 2020;143(5):1315-1331.

63.   Harper L, Fumagalli GG, Barkhof F, et al. MRI visual rating scales in the diagnosis of dementia: Evaluation in 184 post-mortem confirmed cases. *Brain*. 2016;139(4):1211-1225.

64.   Cornblath EJ, Robinson JL, Irwin DJ, et al. Defining and predicting transdiagnostic categories of neurodegenerative disease. *Nature Biomedical Engineering*. 2020;4(8):787-800.

65.   Marquand AF, Wolfers T, Mennes M, Buitelaar J, Beckmann CF. Beyond Lumping and Splitting: A Review of Computational Approaches for Stratifying Psychiatric Disorders. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. 2016;1(5):433-447.

66.   Zhang X, Mormino EC, Sun N, Sperling RA, Sabuncu MR, Yeo BTT. Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. *Proceedings of the National Academy of Sciences of the United States of America*. 2016;113(42):E6535-E6544.

67.   Marquand AF, Rezek I, Buitelaar J, Beckmann CF. Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies. *Biological Psychiatry*. 2016;80(7):552-561.

68.   Marquand AF, Kia SM, Zabihi M, Wolfers T, Buitelaar JK, Beckmann CF. Conceptualizing mental disorders as deviations from normative functioning. *Molecular Psychiatry*. 2019;24(10):1415-1424.

69.   Kia SM, Marquand AF. Neural Processes Mixed-Effect Models for Deep Normative Modeling of Clinical Neuroimaging Data. Published online 2018:1-18.

70.   Kia SM, Huijsdens H, Dinga R, et al. Hierarchical bayesian regression for multi-site normative modeling of neuroimaging data. *arXiv*. Published online 2020.

71.   Ziegler G, Ridgway GR, Dahnke R, Gaser C. Individualized Gaussian process-based prediction and detection of local and global gray matter abnormalities in elderly subjects. *NeuroImage*. 2014;97:333-348.

72.   Huizinga W, Poot DHJ, Vernooij MW, et al. A spatio-temporal reference model of the aging brain. *NeuroImage*. 2018;169:11-22.

73.   Habes M, Pomponio R, Shou H, 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. *Alzheimer's and Dementia*. Published online 2021.

74.   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. *Psychological Medicine*. 2019;50(2):314-323.

75.   Zabihi M, Oldehinkel M, Wolfers T, et al. Dissecting the Heterogeneous Cortical Anatomy of Autism Spectrum Disorder Using Normative Models. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. 2019;4(6):567-578.

76.   Bethlehem RAI, Seidlitz J, Romero-Garcia R, Trakoshis S, Dumas G, Lombardo M V. A normative modelling approach reveals age-atypical cortical thickness in a subgroup of males with autism spectrum disorder. *Communications Biology*. 2020;3(1).

77.   Wolfers T, Doan NT, Kaufmann T, et al. Mapping the Heterogeneous Phenotype of Schizophrenia and Bipolar Disorder Using Normative Models. *JAMA Psychiatry*. 2018;75(11):1146-1155.

78.   Murtagh F, Legendre P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification*. 2014;31(3):274-295.

79.   Cole TJ. The development of growth references and growth charts. *Annals of Human Biology*. 2012;39(5):382-394.

Figure 1 Differences between case-control and data-driven subtype approaches.

(A) The conventional case-control approach. Despite underlying neurobiological heterogeneity, as illustrated by the red, green and blue (RGB) profiles, all patients are analysed together and group average calculated, used to compared against healthy controls to highlight differences between the two groups. Here the average patient (an average of RGB profiles, as circled) assumes neurobiological homogeneity, potentially masking underlying subtypes or individual differences. **(B)** Data driven neuroimaging approach. Cases present neurobiological heterogeneity and are subtyped according to their different neurobiological patterns. This informs the division of the case population into its respective subtypes (distinguished RGB profiles). These subtypes can then inform stratification for further investigation, such as clinical interventions, longitudinal monitoring or genome-wide association studies.

Figure 2 Overview of spatial normative modelling.

(A) The spatial normative model maps centiles of variation across dementia patients and healthy controls. The Gaussian distribution curve (right) illustrates the statistical inference at the level of a dementia patient (red glowing subject) with respect to the normative model. **(B)** Example individual z-score map for the left hemisphere based on cortical thickness. Red colours indicate thinner cortices, relative to the norm, blue thicker cortices. **(C)** Spatial inferences. Spatial normative models are estimated for each sampled brain location in regional space. This can be understood as a set of functions $y = f(x)$, which uses covariates (x) to predict the regional neurobiological variable (y), derived from neuroimaging. **(D)** Procedural summary of spatial normative modelling. Spatial normative model is estimated using a healthy reference cohort. Next this is validated on withheld data (e.g., using cross-validation techniques), to ensure accuracy of the model. The model then can be applied to a dementia cohort. **(E)** Detecting individual trajectories relative to the norm in three different brain regions. Longitudinal data can be used to observe how spatial differences change over time. **(F)** An example of a single spatial trajectory before and after an intervention in a clinical trial. Key: red: dementia group; grey: healthy subjects. For graphs in panels **C-E**, unlabelled axis x: covariates, y: neurobiological variable.
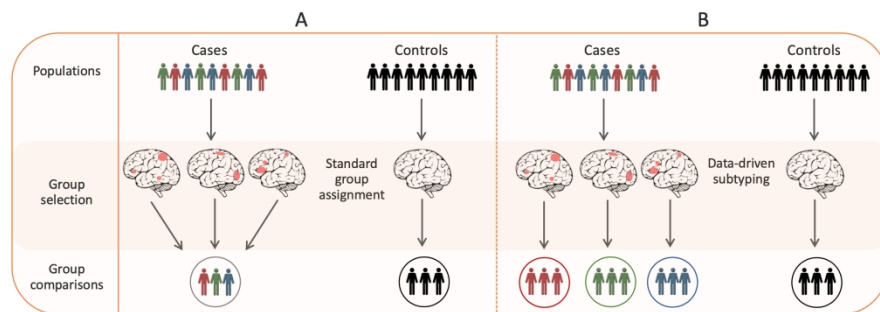
Figure 1 title: Differences between case-control and data-driven subtype approaches

Figure 1 legend:
(A) The conventional case-control approach. Despite underlying neurobiological heterogeneity, as illustrated by the red, green and blue (RGB) profiles, all patients are analysed together and group average calculated, used to compare against healthy controls to highlight differences between the two groups. Here the average patient (an average of RGB profiles, as circled) assumes neurobiological homogeneity, potentially masking underlying subtypes or individual differences. (B) Data-driven neuroimaging approach. Cases present neurobiological heterogeneity and are subtyped according to their different neurobiological patterns. This informs the division of the case population into its respective subtypes (distinguished RGB profiles). These subtypes can then inform stratification for further investigation, such as clinical interventions, longitudinal monitoring or genome-wide association studies.
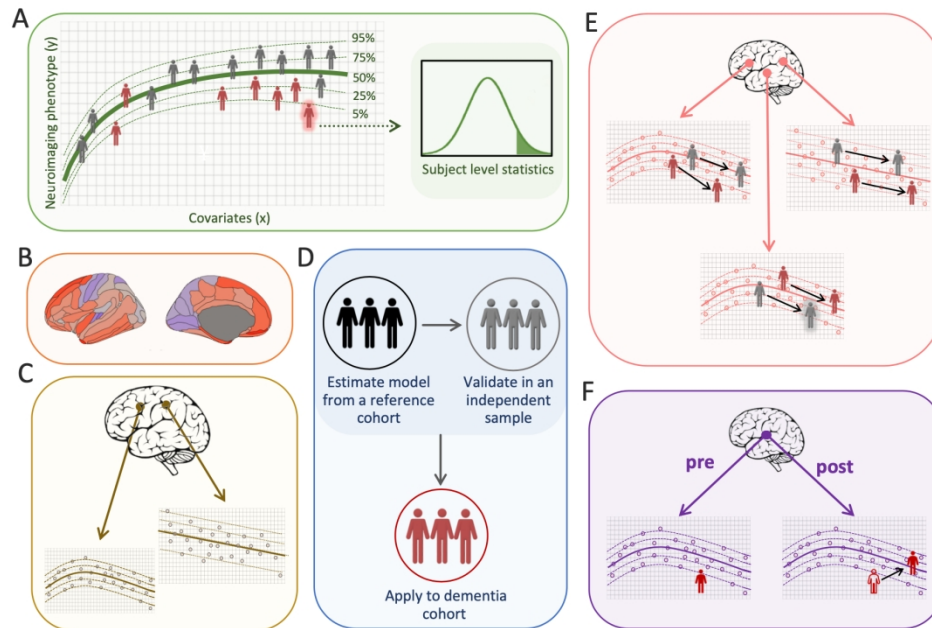
458x203mm (144 x 144 DPI)

Figure 2 title: Overview of spatial normative modelling
Figure 2 legend:
(A) The spatial normative model maps centiles of variation across dementia patients and healthy controls. The Gaussian distribution curve (right) illustrates the statistical inference at the level of a dementia patient (red glowing subject) with respect to the normative model. (B) Example individual z-score map for the left hemisphere based on cortical thickness. Red colours indicate thinner cortices, relative to the norm, blue thicker cortices. (C) Spatial inferences. Spatial normative models are estimated for each sampled brain location in regional space. This can be understood as a set of functions y=f(x), which uses covariates (x) to predict the regional neurobiological variable (y), derived from neuroimaging. (D) Procedural summary of spatial normative modelling. Spatial normative model is estimated using a healthy reference cohort. Next, this is validated on withheld data (e.g., using cross-validation techniques), to ensure the accuracy of the model. The model then can be applied to a dementia cohort. (E) Detecting individual trajectories relative to the norm in three different brain regions. Longitudinal data can be used to observe how spatial differences change over time. (F) An example of a single spatial trajectory before and after an intervention in a clinical trial. Key: red: dementia group; grey: healthy subjects. For graphs in panels C-E, unlabelled axis x: covariates, y: neurobiological variable

341x233mm (144 x 144 DPI)

Box 1: Data-driven statistical methods

Data-driven techniques seek to investigate relationships between the data variables without imposing a priori knowledge of these relationships. Some machine learning techniques can be considered a data-driven approach, where a computer automatically learns (e.g., update model parameters) to optimise performance from experience (i.e., examples of labelled data). This process involves discovering and exploiting regularities in "training data". There are many different problems that can be approached by employing machine learning methods including: anomaly detection, clustering, classification and regression. Broadly speaking, techniques can be summarised into three main groups; supervised, semi-supervised or unsupervised learning. For supervised-learning algorithms, a set of input variables are associated with labels prior to estimating the model. For example, regression analyses (e.g., predicting continuous symptom scores) and classification tasks (e.g., discriminating patients from healthy controls) are examples of supervised learning. Unsupervised learning models the underlying latent structure or distribution in the data to uncover meaningful patterns without supplying a label for each data point.

Clustering is considered unsupervised learning, because the input variable is unlabelled. Clustering aims to identify subtypes, which can be conceptualised as a way to parse a single heterogeneous dataset into a number of more homogeneous subsets **(Fig.1, panel B)**. Models can be derived based on distance, density, connectivity and distribution of the data to be clustered, though such metrics tend to be correlated. Clustering has been the predominant data-driven approach used to explore heterogeneity in dementia.[46] Numerous methods have been implemented, differing in the input features used, the clustering algorithms and the validation approach. Common clustering algorithms include agglomerative, graph-based, and forest-tree based. For example, in agglomerative clustering; the proximity of individual data points (e.g., based on tissue volumes or cortical thickness) are calculated, then similar clusters are merged together to form larger cluster, after which the proximity of new clusters is calculated, these steps are then iterated until all the clusters are merged together to form a single cluster.[78] When implementing these methods is important to consider the type of data used (e.g. dimensionality of data), the choice of algorithm and distance function, the order of the model, the clustering subspace and whether clusters are mutually exclusive (hard clustering), or probabilistic (soft clustering). After implementation, it is also important to assess the number and validity of the clusters generated.[65]

Semi-supervised learning can also be a powerful approach, using clustering techniques alongside supervised classification or regression models to bolster sensitivity in cases of limited training data (e.g., rare diseases). This is where training data points can either be labelled or unlabelled with unlabelled data points aiding the learning of a better classifier (or vice versa). This can sometimes address common confounding effects (e.g., age, sex) which result in clustering the disease effect as transformation from the normal control distribution to the patient distribution, as opposed to just the largest factor of data variability.[25]

Box 2: Spatial normative modelling

Normative models provide statistical inferences at the level of the individual with respect to an expected 'normative' distribution or trajectory over time. This framework is commonly used in growth charts to map developmental changes in body weight and height as a function of age. Deviations from a normal growth curve manifest as outliers from the normative range at each age point.[79] Specifically, spatial normative modelling adopts this concept by modelling the relationship between neurobiological variables (e.g., neuroimaging features which are represented in 'space') and covariates (e.g., demographic variables such as age and sex) to map centiles of variation across a cohort **(Fig.2, Panel A)**. An individual can then be located within the normative distribution to establish to what extent they are an outlier in a given measure. By applying this approach to derive spatial normative models at local brain regions, a map can be generated of where and to what extent an individual's brain differs from the norm **(Fig.2, Panel B)**. Furthermore, by modelling the covariance across the normative cohort at each brain region, confidence intervals can be derived for each point prediction, giving a measure of uncertainty that can be useful for clinical interpretation and subsequent decision-making **(Fig.2, Panel D; E)**.

In the context of dementia, the process to generate these individualised spatial normative brain maps could be as follows: using a separate large reference dataset of healthy participants, spatial normative models of cortical thickness for separate brain regions can be statistically modelled based on age and sex. Next, the parameters of these models would be calibrated using cortical thickness measures derived from a subsample patient cohort under investigation (e.g., dementia patients and scanner-matched controls) From this, z-scores relative to the normative range would be generated for each brain region resulting in a brain 'z-score map' of cortical thickness for each participant in the remaining experimental sample **(Fig.2, Panel B; C)**.

These z-score maps could then be utilised in a variety of research or clinical settings. For example, patients could be 'clustered' based on these neuroanatomical patterns to provide biologically relevant subtypes that may have distinct clinical or biomarker signatures **(Fig.1, Panel B)**. This could provide new mechanistic understanding of dementia as well as facilitate the discovery of genetic influences on dementia-related brain atrophy. Rather than assuming that dementia patients will show common patterns of brain changes, genome-wide association studies could attempt to identify genetic variants that distinguish separate these biologically more homogeneous subtypes from healthy controls. Such subtyping could also be used to stratify enrolment in clinical trials, including only specific subtypes to reduce heterogeneity

and increase power to detect average effects. This could substantially increase sensitivity to treatment effects, reducing the duration and costs of clinical trials.

Going beyond subtyping, the individual patient z-score maps could be used as surrogate outcome measures of treatment efficacy. Rather than simply assessing whether a treatment reduces hippocampal or whole-brain atrophy on average, the magnitude of longitudinal change in z-score maps could be compared between treatment and placebo groups. Importantly, this overcomes assumption of homogeneity, i.e., that a treatment must affect all patients' brains in the same way, slowing atrophy in the same regions. By capturing neuroanatomical heterogeneity at an individual level, spatial normative modelling could indicate whether a treatment slows brain atrophy in different regions in different people, whilst still generating standard effect sizes and confidence intervals for rigorous statistical evaluation.