bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Title page

Accelerated brain ageing and disability in multiple sclerosis

Cole JH*[1,2], Raffel J*[3], Friede T[4], Eshaghi A[5,6], Brownlee W[5], Chard D[5,14], De Stefano N[6], Enzinger C[7], Pirpamer L[8], Filippi M[9], Gasperini C[10], Rocca MA[9], Rovira A[11], Ruggieri S[10], Sastre-Garriga J[12], Stromillo ML[6], Uitdehaag BMJ[13], Vrenken H[14], Barkhof F[5,15,16], Nicholas R*[3,17], Ciccarelli O*[5,16] on behalf of the MAGNIMS study group.

*Contributed equally

[1]Department of Neuroimaging, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK.

[2]Computational, Cognitive and Clinical Neuroimaging Laboratory, Department of Medicine, Imperial College London, London, UK.

[3]Centre for Neuroinflammation and Neurodegeneration, Faculty of Medicine, Imperial College London, London, UK.

[4]Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany.

[5]Queen Square Multiple Sclerosis Centre, Department of Neuroinflammation, UCL Institute of Neurology, London, UK.

[6]Department of Medicine, Surgery and Neuroscience, University of Siena, Siena, Italy.

[7]Research Unit for Neural Repair and Plasticity, Department of Neurology and Division of Neuroradiology, Vascular and Interventional Radiology, Department of Radiology, Medical University of Graz, Graz, Austria.

[8]Neuroimaging Research Unit, Department of Neurology, Medical University of Graz, Graz, Austria

[9]Neuroimaging Research Unit, Institute of Experimental Neurology, Division of Neuroscience, San Raffaele Scientific Institute, Vita-Salute San Raffaele University, Milan, Italy.

[10]Department of Neurosciences, San Camillo-Forlanini Hospital, Rome, Italy.

[11]MR Unit and Section of Neuroradiology, Department of Radiology, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain.

[12]Department of Neurology / Neuroimmunology, Multiple Sclerosis Centre of Catalonia (Cemcat), Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain.

[13]Department of Neurology, VU University Medical Center, Amsterdam, The Netherlands.

[14]Department of Radiology and Nuclear Medicine, Amsterdam University Medical Centers, Amsterdam, The Netherlands.

[15]Centre for Medical Image Computing (CMIC), Department of Computer Science, University College London, UK.

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[16]National Institute for Health Research (NIHR), University College London Hospitals (UCLH) Biomedical Research Centre (BRC), London, UK.

[17]Department of Visual Neuroscience, UCL Institute of Ophthalmology, London, UK.

Address correspondence to Prof Richard Nicholas, Faculty of Medicine, Imperial College London, Charing Cross Campus, Fulham Palace Road, London, W6 8RF, United Kingdom. E-mail: r.nicholas@imperial.ac.uk

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Summary

Background

Brain atrophy occurs in both normal ageing and in multiple sclerosis (MS), but it occurs at a faster rate in MS, where it is the major driver of disability progression. Here, we employed a neuroimaging biomarker of structural brain ageing to explore how MS influences the brain ageing process.

Methods

In a longitudinal, multi-centre sample of 3,565 MRI scans in 1,204 MS/clinically isolated syndrome (CIS) patients and 150 healthy controls (HCs) (mean follow-up time: patients 3·41 years, HCs 1·97 years) we measured 'brain-predicted age' using T1-weighted MRI. Brain-predicted age difference (brain-PAD) was calculated as the difference between the brain-predicted age and chronological age. Positive brain-PAD indicates a brain appears older than its chronological age. We compared brain-PAD between MS/CIS patients and HCs, and between disease subtypes. In patients, the relationship between brain-PAD and Expanded Disability Status Scale (EDSS) at study entry and over time was explored.

Findings

Adjusted for age, sex, intracranial volume, cohort and scanner effects MS/CIS patients had markedly older-appearing brains than HCs (mean brain-PAD 11·8 years [95% CI 9·1—14·5] versus -0·01 [-3·0—3·0], p<0·0001). All MS subtypes had greater brain-PAD scores than HCs, with the oldest-appearing brains in secondary-progressive MS (mean brain-PAD 18·0 years [15·4—20·5], p<0·05). At baseline, higher brain-PAD was associated with a higher EDSS, longer time since diagnosis and a younger age at diagnosis. Brain-PAD at study entry significantly predicted time-to-EDSS progression (hazard ratio 1·02 [1·01—1·03], p<0·0001): for every 5 years of additional brain-PAD, the risk of progression increased by 14·2%.

Interpretation

MS increases brain ageing across all MS subtypes. An older-appearing brain at baseline was associated with more rapid disability progression, suggesting 'brain-age' could be an individualised prognostic biomarker from a single, cross-sectional assessment.

Funding

UK MS Society; National Institute for Health Research University College London Hospitals Biomedical Research Centre.

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Research in context

Evidence before this study

We searched Pubmed and Scopus with the terms “multiple sclerosis” and “brain ageing” or “brain age” and “neuroimaging” or “MRI” for studies published before 15th March 2019. This searched return no studies of brain ageing in multiple sclerosis. We also searched the pre-print server for biology, bioRxiv, and found one manuscript deposited, though this study has yet to appear in a peer-reviewed journal. This study found a strong effect of multiple sclerosis on the apparent age of the brain, though was only cross-sectional, was from a single centre, did not consider disease subtypes and did not consider the relevance of clinical characteristics for brain ageing. Therefore, although there is strong prior evidence of the importance of brain atrophy in multiple sclerosis, there was no information on how the nature of this atrophy relates to brain ageing.

Added value of this study

Here we demonstrate for the first time that the progressive atrophy in multiple sclerosis patients results in an acceleration of age-related changes to brain structure. Using a large multi-centre study, our data strongly support the idea that brain ageing is increased in multiple sclerosis, and that this is apparent across disease subtypes, including those with very early disease - Clinically Isolated Syndrome. Of particular value is the demonstration that baseline brain-age can be used to predict future worsening of disability, suggesting that a general index of age-related brain health could have relevance in clinical practice for predicting which patients will go on to experience a more rapidly progressing disease course.

Implications of all the available evidence

Combined with the single other available study, this work shows robust evidence for a cross-sectional influence of multiple sclerosis on the apparent age of the brain, under the brain-age paradigm. This paradigm provides a new approach to considering how multiple sclerosis effects the structure of the brain during ageing, suggesting that multiple sclerosis may result in both disease-specific insults (e.g., lesions) alongside changes that are less specific (e.g., atrophy) and seen in ageing and other diseases. Potentially, treatments that improve brain health during normal ageing could be used to benefit patients with multiple sclerosis. Finally, brain-age may also have prognostic clinical value as a sensitive, if non-specific, biomarker of future health outcomes.

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

In multiple sclerosis (MS), age has been implicated as the dominant driver of disease progression.[1] Older age increases the risk of progression,[2] irrespective of disease duration, and once progression starts, disability accrual is independent of previous evolution of the disease (presence or not relapses, or relapse rates.[3-5] This raises the possibility that MS interacts with some of the neurobiological drivers of brain ageing, leading to acceleration of the process, hastening brain atrophy in some individuals and leading to poorer long-term outcomes.[6]

That diseases may impact rates of biological ageing has been previously mooted outside of the context of MS. Potentially, a disease has both a specific impact but also may trigger a sequence of events which result in an acceleration of the biological processes seen in normal ageing, both systemically[7] and in the brain.[8,9]

Recently, methods have been developed for measuring the biological ageing of the brain; the so-called 'brain-age' paradigm.[10] Brain-age uses machine-learning analysis to generate a prediction of an individual's age (their brain-predicted age), based solely on neuroimaging data (most commonly 3D T1-weighted MRI). The comparison of an individual's brain-predicted age with their chronological age thus gives an index of whether their brain structure appears 'older' or 'younger' than would be expected for their age. By subtracting chronological age from brain-predicted age one can derive a brain-predicted age difference (brain-PAD); a simple numerical value in the unit years which shows promise as a biomarker of brain ageing. For example, brain-age has been shown to predict the likelihood of conversion from mild cognitive impairment to Alzheimer's[11,12] as well as the risk of mortality.[13] Moreover, there is evidence for increased brain ageing in other neurological conditions contexts: traumatic brain injury,[14] HIV,[15] Down's syndrome,[16] and epilepsy.[17]

Here we employ brain-age to assess the relationship between MS disease progression and the brain ageing process. Using longitudinal neuroimaging and clinical outcomes in a large cohort of MS patients and healthy controls (HCs), we tested the following hypotheses: (i) MS patients have older-appearing brains than HCs; (ii) In MS patients, there is a relationship between brain-predicted age difference and disability at study entry; (iii) Brain-predicted age difference increases over time as disabilities worsen; and (iv) Brain-predicted age difference at baseline predicts future disability progression.

Methods

Participants

This study used data collected from seven European MS centres (MAGNIMS: www.magnims.eu) and Imperial College London on n=1,354 participants (table 1), largely overlapping with our previous work (detailed in the appendix table S1).[18] Patients had all received a diagnosis of MS according to 2010 McDonald Criteria[19] or CIS.[20] MS/CIS patients were scored on the Expanded Disability Status

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Scale (EDSS).[21] HCs without history of neurological or psychiatric disorders were also included (n=150). For longitudinal imaging analysis, participants were required to have undergone at least two high-resolution T1-weighted MRI acquired with the same protocol with an interval of ≥1 month.

The final protocol for this study was reviewed and approved by the European MAGNIMS collaboration for analysis of pseudo-anonymized scans and the Imperial NHS Trust (London Riverside Research Ethics Committee: 14/LO/0343).

EDSS progression

Time-to-event, where a progression event was an individual's progression on the EDSS, was defined as per our previous work[18]: when a patient showed a longitudinal change of: a 1·5-point increase in EDSS if the baseline EDSS was 0; a 1-point increase if baseline EDSS was 1 to 6 inclusive; and a 0·5-point increase if EDSS was greater than 6.

Neuroimaging acquisition

Overall, 3,565 T1-weighted MRI scans were used in the study according to local MRI protocols, which used similar acquisition parameters. Thirteen different scanners (Siemens, GE, Philips) were used in patients recruited from 1998 onwards (see appendix table S1).

Machine-learning brain-predicted age analysis

Brain-predicted age calculation followed our previously established protocol.[15] In brief, all structural images were pre-processed using SPM12 to generate grey matter (GM), white matter (WM) segmentations. Visual quality control was then conducted to verify segmentation accuracy; all images were included. Segmented GM and WM images were then non-linearly registered to a custom template (based on the training dataset). Finally, images were affine registered to MNI152 space (voxel size = 1·5mm^3), modulated and smoothed (4mm). Summary volumetric measures of GM, WM, cerebrospinal fluid (CSF) and intracranial volume (ICV) were also generated.

Brain-predicted ages were generated using Pattern Recognition for Neuroimaging Toolbox (PRoNTo v2·0, www.mlnl.cs.ucl.ac.uk/pronto) software.[22] First, a model of healthy brain ageing was defined: brain volumetric data (from in a separate training dataset, n=2001 healthy people, aged 18-90; appendix table S2) were used as the independent variables in a Gaussian Processes regression, with age as the dependent variable. This regression model achieved a mean absolute error (MAE) of 5·02 years, assessed using ten-fold cross-validation, which explained 88% of the variance in chronological age.

Next, the coefficients from the full historical training model (n=2001) were applied to the current test data (i.e., MS/CIS patients and HCs), to generate brain-predicted ages. These values were adjusted to remove age-related variance, by subtracting 3·33 and then dividing by 0·91 (the intercept and slope of a linear regression of brain-predicted age on chronological age in the training dataset).

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Finally, brain-PAD scores were calculated by subtracting chronological age from brain-predicted age and used for subsequent analysis. A positive brain-PAD score indicates that the individual's brain is predicted to be 'older' than their chronological age.

Statistical analysis

Using brain-PAD values further statistical analysis was carried out to test our hypotheses, using R v3·4·3. A full list of R packages and versions is included in the accompanying R Notebook (appendix). We used linear mixed effects models, enabling incorporation of fixed and random effects predictors to model each given outcome measure. In these models, brain-PAD was used as the outcome variable. Each model included fixed effects of group (e.g., MS/CIS patient versus HCs; MS subtype [CIS, RRMS, SPMS, PPMS]), age, sex and ICV and random effects of MRI scanner field-strength and original study cohort (modelling intercept). Estimated marginal means and confidence intervals from linear models were calculated. This analysis was repeated using data from a single cohort from a single centre (UCL, London), where all MS subtypes were present.

A random effects meta-analysis was conducted to explore the heterogeneity of the group effects on brain-PAD across different study cohorts. Only cohorts that included HCs and MS or CIS patients were included in this analysis.

To establish whether brain volume measurements were driving the variability in brain-PAD, we performed a linear regression with hierarchical partitioning of variance, with brain-PAD as the outcome variable and age, sex, GM, WM and CSF volume as predictors.

Subsequent analyses were conducted to test for fixed-effect influences of EDSS score (MS and CIS patients), and time since clinical diagnosis and age at clinical diagnosis (MS patients only). Model fits were considered using F-tests and post-hoc pairwise comparisons using t-tests or Tukey tests where appropriate.

We explored how longitudinal changes in brain-PAD related to changes in disability over time in two ways: (i) by correlating annualised change in brain-PAD (i.e., the difference between first measured brain-PAD and last brain-PAD, divided by the interval in years) with the annualised change in EDSS score; (ii) by using linear mixed effects models to investigate group (MS/CIS vs., HCs; MS subtype) by time interactions. These analyses included a random effect of participant (modelling slope and intercept), alongside age, sex, ICV scanner and cohort effects.

Survival analysis, using a Cox proportional hazards regression, was used to test whether baseline brain-PAD predicted time-to-EDSS progression, including age at baseline MRI and sex as covariates.

We investigated the impact of MS lesions on brain-PAD in MS. Using cross-sectional data from a subset of n=575 MS/CIS patients, for which manually-annotated lesion maps were available, we

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

explored the relationship between MS lesions and measurements of brain-PAD, using the FSL lesion-filling algorithm,[23] by artificially removing lesions from T1-weighted MRI scans. Both 'lesion-filled' and 'unfilled' scans were run through the brain-age prediction procedure, then resulting brain-PAD scores compared.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Multiple sclerosis is associated with older appearing brains

The MAGNIMS sample forms part of a well-characterised population (table 1). The combined cohort involves patients from six countries with a mean follow-up of 3·41 years in patients.

Patients with MS/CIS had markedly greater brain-PAD scores at time of initial MRI scan compared to HCs (estimated marginal means 11·8 years, [95% CI 9·1–14·5] versus -0·01 [95% CI -3·0–3·0]). When adjusted for the age, sex, intracranial volume, cohort and scanner effects, there was a statistically significant group mean difference in brain-PAD of 11·8 years (95% CI 9·9–13·8, $p<0·0001$).

Though there is considerable heterogeneity between the study cohorts, due to clinical characteristics and technical factors (e.g., MRI scanner system), the difference between MS/CIS and HCs was robust in a random-effects meta-analysis of a subset of the data; six London cohorts that included both MS/CIS patients and HCs (figure 1A). The heterogeneity in the group differences was substantial (I^2 = 59%, [95% CI 3–91%]).

MS subtype (CIS, RRMS, SPMS, PPMS) significantly influenced brain-PAD ($F_{3,802·25}$ = 29·9, $p<0·0001$, figure 1B). Estimated marginal mean brain-PAD per subtype were: CIS 6·3 years [95% CI 3·9–8·8], RRMS 12·4 years [95% CI 10·3–14·5], SPMS 18·0 years [95% CI 15·4–20·5], and PPMS 12·4 years [95% CI 9·7–15·2]. Post-hoc pairwise group comparison (appendix table S3) showed statistically significant differences ($p<0·05$) in brain-PAD between each subtype and HCs, and between CIS patients and each of the three MS groups (RRMS, SPMS, PPMS). SPMS patients showed significantly greater brain-PAD compared to both RRMS and PPMS patients. The difference in brain-PAD between PPMS and RRMS was not statistically significant ($p=0·62$). The findings of differences in brain-PAD between MS subtypes were replicated in a single cohort from a single centre, where all subtypes were present (cohort UCL3, figure 1C). Brain-PAD scores and corresponding T1-weighted MRI scans of individual female participants with similar ages, but with different subtypes of MS, are illustrated in figure 1D.

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The relationship between lesions, brain volume and brain-PAD

We considered the impact of lesions of brain-PAD, by comparing brain-PAD values on a single MRI scan from n=575 patients with both a lesion-filled and unfilled version of the same image. The correlation between brain-predicted age using filled and unfilled scans was r=0·99, p<0·0001 (appendix figure S1A) suggesting that the presence of lesions did not overly influence the brain-PAD values used throughout the study (which were unfilled). A Bland-Altman plot showed a mean difference between filled and unfilled scans was -0·28 ±1·29 years with no systemic bias caused by lesion filling evident, though there was increased variability between ages 60-70 years (appendix figure S1B).

When we examined whether brain volume measurements were driving the variability in brain-PAD, we found that the combination of chronological age, sex, GM, WM and CSF volume explained about half of the variation in brain-PAD (adjusted R^2=0·48) (appendix table S4). Age (9% variance explained), GM (15%) and CSF (20%) volume were major contributors to variance in brain-PAD.

Brain-PAD at baseline is associated with disability, age at diagnosis, and time since clinical diagnosis

At baseline, a higher brain-PAD was associated with higher disability, as measured by the EDSS, when adjusting for age, sex, ICV, scanner and cohort: for every 1·74 years increase in brain-PAD, the EDSS increased by one point (95% CI 1·39–2·09], p<0·0001). This effect was consistent across the MS subtypes with no statistically significant interaction between subtype and EDSS score ($F_{3,1159·6}$ = 1·12, p=0·34; figure 2A). With the same adjustments, a higher brain-PAD was associated with both younger age at diagnosis and longer time since diagnosis: for every year increase in brain-PAD, the age at diagnosis was reduced by 0·45 years (95% CI -0·55–-0·36], p<0·0001); for every 0·48 year increase in brain-PAD, the time since diagnosis increased by one year (95% CI 0·40–0·57, p<0·0001). There was an interaction between subtype (RRMS, PPMS and SPMS) and age at diagnosis ($F_{2,883·9}$ = 3·20, p=0·041; figure 2B), driven by the presence of stronger relationships between brain-PAD and age at diagnosis in PPMS (slope beta -0·51) and SPMS (beta -0·57) compared to RRMS (beta -0·36), though all were significant (p<0·001). For time since diagnosis, the interaction was also significant ($F_{2,690·5}$ = 3·61, p=0·028; figure 2C), driven by the presence of relationships in RRMS (beta 0·48, p<0·0001) and SPMS (beta 0·26, p=0·01), not observed in PPMS (beta 0·12, p=0·47).

Brain-PAD increase over time correlates with EDSS worsening

In patients who had two or more scans (n=1155), there was a significant positive correlation between annualised change in brain-PAD and annualised change in EDSS (Pearson's r=0·26, p<0·0001). There was a significant interaction between EDSS change and disease subtype, when predicting brain-PAD change in linear model ($F_{4,1092}$ = 24·5, p=0·009). The slopes were positive in CIS (beta 0·84, p=0·0001) and RRMS (beta 1·25, p<0·0001), though flatter in PPMS (beta 0·59,

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

p=0·090) and negative (though not significant) in SPMS (beta -0·70, p=0·29; figure 3). To explore the latter finding post-hoc, correlated baseline brain-PAD with the number of follow-up scans completed. This showed a significant inverse correlation (n=104, Spearman's rho=-0·29, p=0·0028).

Brain-predicted age difference at first scan predicts EDSS worsening

In patients who had EDSS assessed at ≥2 time-points (n=1147), baseline brain-PAD significantly predicted EDSS worsening. Of these patients, 303 (26·5%) experienced EDSS worsening during the follow-up period. Using a Cox proportional-hazards regression model, adjusted for age and sex, the hazard ratio for brain-PAD was 1·027 (95% CI 1·016–1·038, p<0·0001). In other words, for every 5 years of additional brain-PAD, there was a 14·1% increased chance of EDSS progression during follow-up. Survival curves grouped by a median split of baseline brain-PAD illustrate the differing rates of 'survival' prior to EDSS progression (figure 4).

MS accelerates longitudinal increase in brain-PAD

A total of 1266 participants had two or more MRI scans (MS/CIS=1155, HCs=111). This included 573 with three or more scans (MS/CIS=509, HCs=64). When using these data, we found a significant interaction between group and time ($F_{1,1325\cdot6}$ = 5·37, p=0·021) and between MS subtypes and time ($F_{4,938\cdot25}$ = 5·35, p<0·0001), when adjusting for age, gender, ICV, cohort and scanner status (figure 5). This indicated that the annual rate of increase in brain-PAD over time was faster in MS/CIS than in HCs, and significantly different between MS subtypes. The estimate marginal mean annualised rates of increase in brain-PAD per group was as follows: HCs -0·98 [95% CI -2·03–0·07], CIS -0·14 [95% CI -1·07–0·78], RRMS 0·93 [95% CI 0·21–1·66], SPMS 0·34 [95% CI -0·69–1·37], PPMS 1·21 [95% CI 0·16–2·25], all CIS/MIS 0·70 [95% CI 0·01–1·39].

Discussion

By assessing the relationship between MS disease progression and the normal brain ageing process, we have found that patients with MS have an older appearing brain (i.e., higher brain-PAD) compared to controls. As the disease develops from a clinically isolated episode to relapsing and then secondary progressive MS, brain-PAD increases. A single baseline brain-PAD was independently associated with higher disability (measured by EDSS), younger age at diagnosis and longer time since diagnosis, irrespectively of disease phenotype. Using scans performed at multiple sites in different scanners we observed that longitudinal brain-PAD increases correlate with worsening disability and that measures of brain-PAD at baseline predict future disability accumulation. In the whole cohort, we show that measures of brain-PAD over time increase with respect to chronological age, implying an accelerated ageing process, particularly in RRMS and PPMS.

In a life-long disease, the accumulation of neurological disability is the main clinical and societal burden,[24] estimated to cost $10·6 billion/year in the USA.[25] Tracking disease evolution is hampered

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

by the lack of a simple and powerful outcome measure. MRI-assessed brain atrophy is a surrogate outcome for this process, but the need for precise longitudinal assessments, usually over at least a 12-month interval, reduces the feasibility of use. Here, we demonstrate that with a single T1-weighted MRI, brain-PAD values can index elements of MS disease progression. Firstly, we show that a single point estimate can place a patient's disease and disability in context of their age. This has been lacking with current techniques but is achieved because brain-PAD measures change relative to a model of the healthy ageing process. Our results suggest that the 'brain-age' framework can provide informative data without the need for longitudinal scans.[26] Secondly, we demonstrate that a single measure can give prognostic value for disability accumulation. This can allow us to better contextualise the impact of the disease on an individual, measured at a single time point, and then chart different pathways of neurodegeneration in MS. Brain atrophy has undoubted utility in capturing elements of disease progression, but is currently difficult to utilise in clinical practice.[27] Here we demonstrate that machine learning technique provides a biomarker of structural brain ageing that enables prediction of disability worsening. Thus, the ability to make prognostic predictions from cross-sectional data could prove highly valuable to facilitate early use of therapy to prevent future disability accumulation.[28]

The brain-age paradigm has been applied widely in neuropsychiatric diseases,[10] though only recently in MS.[29] Kaufmann and colleague's analysis (n=254) showed a strong effect of MS on brain-age (mean increase 5·6 years), though was only cross-sectional and did not explore subtypes separately. Here we go further, utilising serial MRI scans that were carried out over 15 years in a wide range of settings – different countries, institutions and scanners. The mean magnitude of the apparent brain ageing we observed MS (11.8 years) is greater than has been reported in dementia (9 years),[11] epilepsy (4·5 years)[17] or after a traumatic brain injury (4·7 years).[14] We show that brain-PAD increases faster than chronological age in MS/CIS patients, suggesting an accelerating neurodegenerative process. Interestingly, brain-PAD did not increase longitudinally in SPMS patients; potentially due to a survivor bias or a floor effect in this group, whereby those patients with rapidly deteriorating disease did not return for longitudinal follow-up. Evidence for this comes from the inverse correlation between brain-PAD at baseline and the number of follow-up scans acquired in SPMS patients.

We addressed some potential issues with the use of a non-specific ageing biomarker like brain age for the assessment of MS. Brain lesions, the overt MRI marker of MS disease activity, had minimal impact of the brain-PAD measurement in MS. Brain volumes, perhaps unsurprisingly, were strongly correlated with brain-PAD; GM, WM and CSF volume measures combined explained ~49% of the variance in brain-PAD. Evidently, a substantial proportion of variation in brain-PAD is not explained by demographic and MRI characteristics and might be unique to 'brain-age'. In particular, 'brain-age' incorporates voxelwise MRI data in the statistical model, thereby capturing more information than

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

when using summary statistics. This means that more widespread and distributed patterns of features (i.e., voxelwise GM and WM volumes) can contribute to the age-prediction model, capturing elements of cortical thinning, sulcal widening and ventricular enlargement, alongside more macroscopic loss of tissue volume.

Our study has some strengths and weaknesses. The sample size for both training and test sets is relatively large but one potential limitation is the multiple sources of training data, though previous work has shown high between-scanner reliability.[30] Thus, if it is to be used as a single value this would need to be in the context of individual scanner performance. Comprehensive biomedical data were not available on all these individuals in the training dataset, meaning some may have had undetected health conditions. However, individuals in this sample were screened according to various criteria to ensure the absence of manifest neurological, psychiatric or major medical health issues. We were not able to determine the impact of therapy in this study as it was not a randomised trial and worsening disease drives use of therapy, the effectiveness of which is challenging to determine. However, the majority of the current study sample were on not receiving therapy at baseline, thus therapeutic effects are unlikely to have confounded our results.

This work supports the use of the 'brain-age' paradigm in MS. We propose that brain-predicted age has potential value for: 1) MS disease monitoring; potentially capturing the progressive processes that start early on in all disease phenotypes including CIS. 2) Integrating MRI measures of brain injury in MS in a wide range of centres and different scanners. 3) Conveying complex neuroanatomical information in a conceptually simple and intuitive manner. 4) Assessing both current brain health and prognosis. 5) Aiding clinical trial design, by stratifying enrolment based on high brain-PAD, or using brain-PAD as a surrogate outcome measure, reflecting age-associated neurodegeneration. Further work is needed to determine its utility in larger clinical cohorts, but its ease of use makes it an exciting candidate for such cohorts. Further work is needed to improve the anatomical interpretability of brain-age, both in general and specifically to MS. Ultimately, this may offer insight into an individual's disease course, in line with the move towards precision medicine in the treatment of MS.

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements

JC is funded by a UKRI/MRC Innovation Fellowship. OC, RN, FB and DC acknowledge the National Institute for Health Research University College London Hospitals Biomedical Research Centre. RN acknowledges the National Institute for Health Research Imperial College London Hospitals Biomedical Research Centre. AE received the McDonald Fellowship from Multiple Sclerosis International Federation (MSIF, http://www.msif.org) and ECTRIMS-MAGNIMS fellowship.

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Leray E, Yaouanq J, Le Page E, et al. Evidence for a two-stage disability progression in multiple sclerosis. *Brain : a journal of neurology* 2010; **133**(Pt 7): 1900-13.

2. Scalfari A, Lederer C, Daumer M, Nicholas R, Ebers GC, Muraro PA. The relationship of age with the clinical phenotype in multiple sclerosis. *Mult Scler* 2016; **22**(13): 1750-8.

3. Koch M, Mostert J, Heersema D, De Keyser J. Progression in multiple sclerosis: further evidence of an age dependent process. *J Neurol Sci* 2007; **255**(1-2): 35-41.

4. Scalfari A, Neuhaus A, Daumer M, Ebers GC, Muraro PA. Age and disability accumulation in multiple sclerosis. *Neurology* 2011; **77**(13): 1246-52.

5. Tutuncu M, Tang J, Zeid NA, et al. Onset of progressive phase is an age-dependent clinical milestone in multiple sclerosis. *Mult Scler* 2013; **19**(2): 188-98.

6. Oost W, Talma N, Meilof JF, Laman JD. Targeting senescence to delay progression of multiple sclerosis. *J Mol Med (Berl)* 2018; **96**(11): 1153-66.

7. Franceschi C, Garagnani P, Morsiani C, et al. The Continuum of Aging and Age-Related Diseases: Common Mechanisms but Different Rates. *Frontiers in Medicine* 2018; **5**(61).

8. Mattson MP, Arumugam TV. Hallmarks of Brain Aging: Adaptive and Pathological Modification by Metabolic States. *Cell Metabolism* 2018; **27**(6): 1176-99.

9. Cole JH. Neuroimaging Studies Illustrate the Commonalities Between Ageing and Brain Diseases. *BioEssays* 2018; **40**(7): 1700221.

10. Cole JH, Franke K. Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers. *Trends in neurosciences* 2017; **40 12**: 681-90.

11. Franke K, Gaser C. Longitudinal changes in individual BrainAGE in healthy aging, mild cognitive impairment, and Alzheimer's Disease. *GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry* 2012; **25**(4): 235-45.

12. Gaser C, Franke K, Klöppel S, Koutsouleris N, Sauer H, for the Alzheimer's Disease Neuroimaging Initiative. BrainAGE in mild cognitive impaired patients: predicting the conversion to Alzheimer's disease. *PloS one* 2013; **8**(6): e67346.

13. Cole JH, Ritchie SJ, Bastin ME, et al. Brain age predicts mortality. *Molecular psychiatry* 2018; **23**: 1385-92.

14. Cole JH, Leech R, Sharp DJ, for the Alzheimer's Disease Neuroimaging Initiative. Prediction of brain age suggests accelerated atrophy after traumatic brain injury. *Ann Neurol* 2015; **77**(4): 571-81.

15. Cole JH, Underwood J, Caan MWA, et al. Increased brain-predicted aging in treated HIV disease. *Neurology* 2017; **88**(14): 1349-57.

16. Cole JH, Annus T, Wilson LR, et al. Brain-predicted age in Down Syndrome is associated with β-amyloid deposition and cognitive decline. *Neurobiology of aging* 2017; **56**: 41-9.

17. Pardoe HR, Cole JH, Blackmon K, Thesen T, Kuzniecky R. Structural brain changes in medically refractory focal epilepsy resemble premature brain aging. *Epilepsy Research* 2017; **133**: 28-32.

18. Eshaghi A, Prados F, Brownlee Wallace J, et al. Deep gray matter volume loss drives disability worsening in multiple sclerosis. *Ann Neurol* 2018; **83**(2): 210-22.

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

19. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol* 2011; **69**(2): 292-302.

20. Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology* 2014; **83**(3): 278-86.

21. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983; **33**(11): 1444-52.

22. Schrouff J, Rosa MJ, Rondina JM, et al. PRoNTo: Pattern recognition for neuroimaging toolbox. *Neuroinformatics* 2013; **11**(3): 319-37.

23. Battaglini M, Jenkinson M, De Stefano N. Evaluating and reducing the impact of white matter lesions on brain volume measurements. *Hum Brain Mapp* 2012; **33**(9): 2062-71.

24. Degenhardt A, Ramagopalan SV, Scalfari A, Ebers GC. Clinical prognostic factors in multiple sclerosis: a natural history review. *Nature reviews Neurology* 2009; **5**(12): 672-82.

25. Gooch CL, Pracht E, Borenstein AR. The burden of neurological disease in the United States: A summary report and call to action. *Ann Neurol* 2017; **81**(4): 479-84.

26. Uher T, Vaneckova M, Krasensky J, et al. Pathological cut-offs of global and regional brain volume loss in multiple sclerosis. *Mult Scler* 2017: 1352458517742739.

27. Rocca MA, Battaglini M, Benedict RH, et al. Brain MRI atrophy quantification in MS: From methods to clinical application. *Neurology* 2017; **88**(4): 403-13.

28. Montalban X, Gold R, Thompson AJ, et al. ECTRIMS/EAN Guideline on the pharmacological treatment of people with multiple sclerosis. *Mult Scler* 2018; **24**(2): 96-120.

29. Kaufmann T, van der Meer D, Doan NT, et al. Genetics of brain age suggest an overlap with common brain disorders. *bioRxiv* 2018.

30. Cole JH, Poudel RPK, Tsagkrasoulis D, et al. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage* 2017; **163C**: 115-24.

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tables

Table 1. Characteristics of MS/CIS patients and healthy controls.

	MS/CIS patients	Healthy controls
N	1204	150
N with follow-up data	1155	111
Female (n, %)	771 (64)	82 (55)
Number of scans per participant mean ± SD [range]	2·61 ± 1·01 [1-7]	2·82 ± 1·90 [1-10]
Length of follow-up (years) mean ± SD [range]	3·41 ± 3·15 [0·2-15]	1·97 ± 1·38 [0·5-6·0]
Age at baseline scan (years)	39·41 ± 10·76 [15-68]	37·29 ± 9·96 [23-66]
Brain-predicted age at baseline (years)	50·27 ± 14·90 [7·4-92]	38·43 ± 11·12 [14·5-70]
Baseline subtype (%): CIS/RRMS/SPMM/PPMS	296 (25) / 677 (56) / 111 (9) / 120 (10)	-
Time since **clinical** diagnosis at baseline (years)	7·26 ± 7·96 [0-48]	-
EDSS at baseline mean ± SD [range]	2·60 ± 1·95 [0-9]	-
Disease-modifying treatment at baseline (n, %): yes/no/unknown	475 (39) / 675 (56) / 54 (5)	-

CIS = clinically isolated syndrome, EDSS = Expanded Disability Status Scale, RRMS = relapsing remitting MS, SD = standard deviation, SPMS = secondary progressive MS, PPMS = primary progressive MS.

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figures

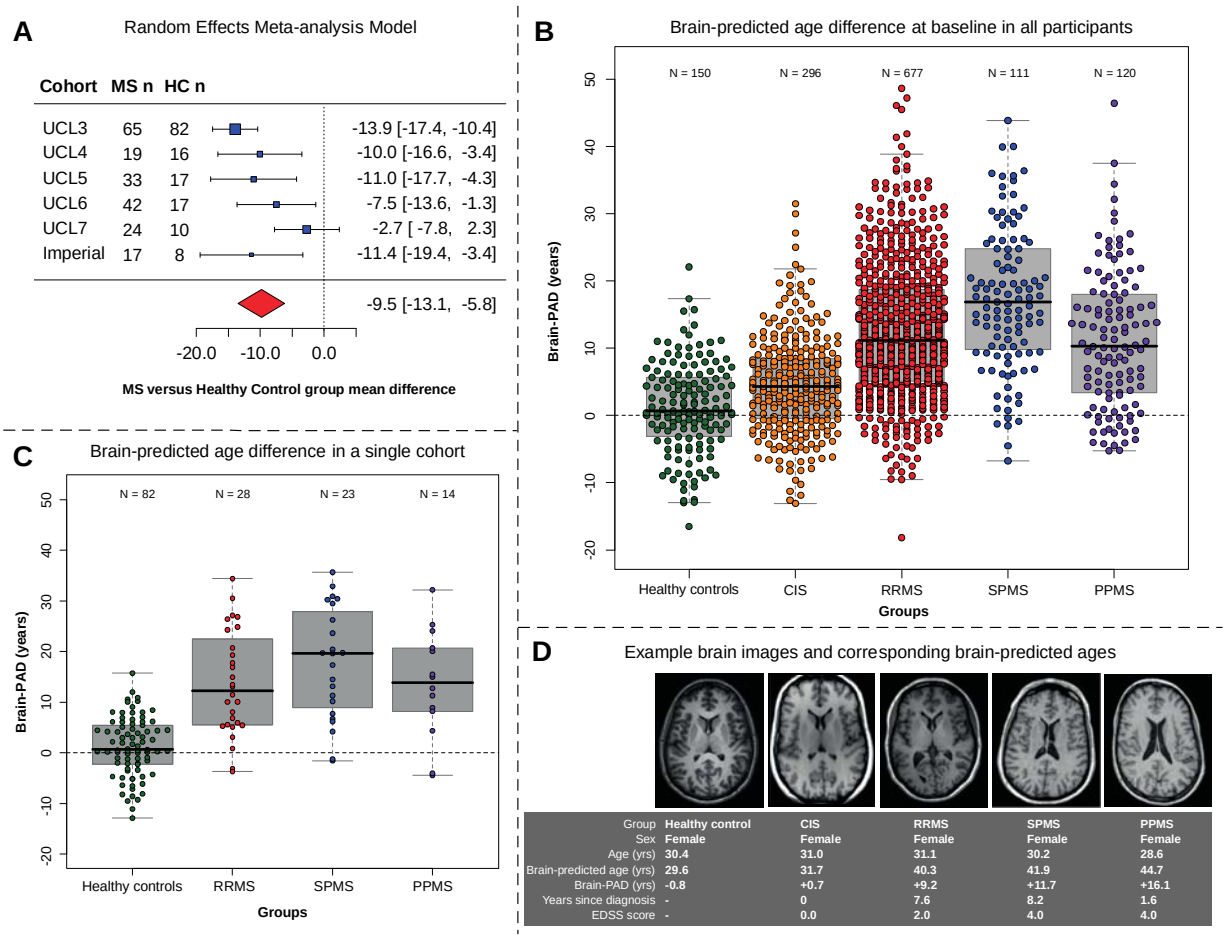


Figure 1. Brain-predicted age difference (Brain-PAD) for MS/CIS patients and healthy controls at baseline

A) A random-effects meta-analysis of the six cohorts that included both MS/CIS patients and HCs found the pooled effect of MS/CIS on brain-PAD compared to HCs was 9·45 years (95% CI 13·11–5·80), across a total of n=200 MS/CIS patients and n=15HC. Heterogeneity was estimated at I^2 = 59% [3–91%]. B) Grouped scatterplot depicting the distributions of brain-PAD at baseline, in years. Black lines represent the group median, shaded boxes show the inter-quartile range and whiskers 1·5 times the inter-quartile range from the median. C) Data from cohort "UCL3", where all MS subtypes were present, confirms a similar result to the total cohort. D) Examples of how brain structure relates to brain-PAD, with axial slice from T1-weighted MRI from one healthy control and four individuals with CIS or MS, all females of a similar age. A control brain from a 30-year-old female with a brain-PAD of -0·8 years can be compared to a 31-year-old female with CIS, EDSS of 0·0 and a brain-PAD of +0·7 years, and 31-year-old with RRMS, EDSS of 2·0 and a brain-PAD of +9·2 years. In addition, we illustrate a 30-year-old with SPMS, EDSS of 4·0 and a brain-PAD of +11·7 years and a 28-year-old with PPMS, EDSS of 4·0 and a brain-PAD of +16·7 years. CIS = clinically isolated syndrome, RRMS = relapsing remitting MS, SPMS = secondary progressive MS, PPMS = primary progressive MS.

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

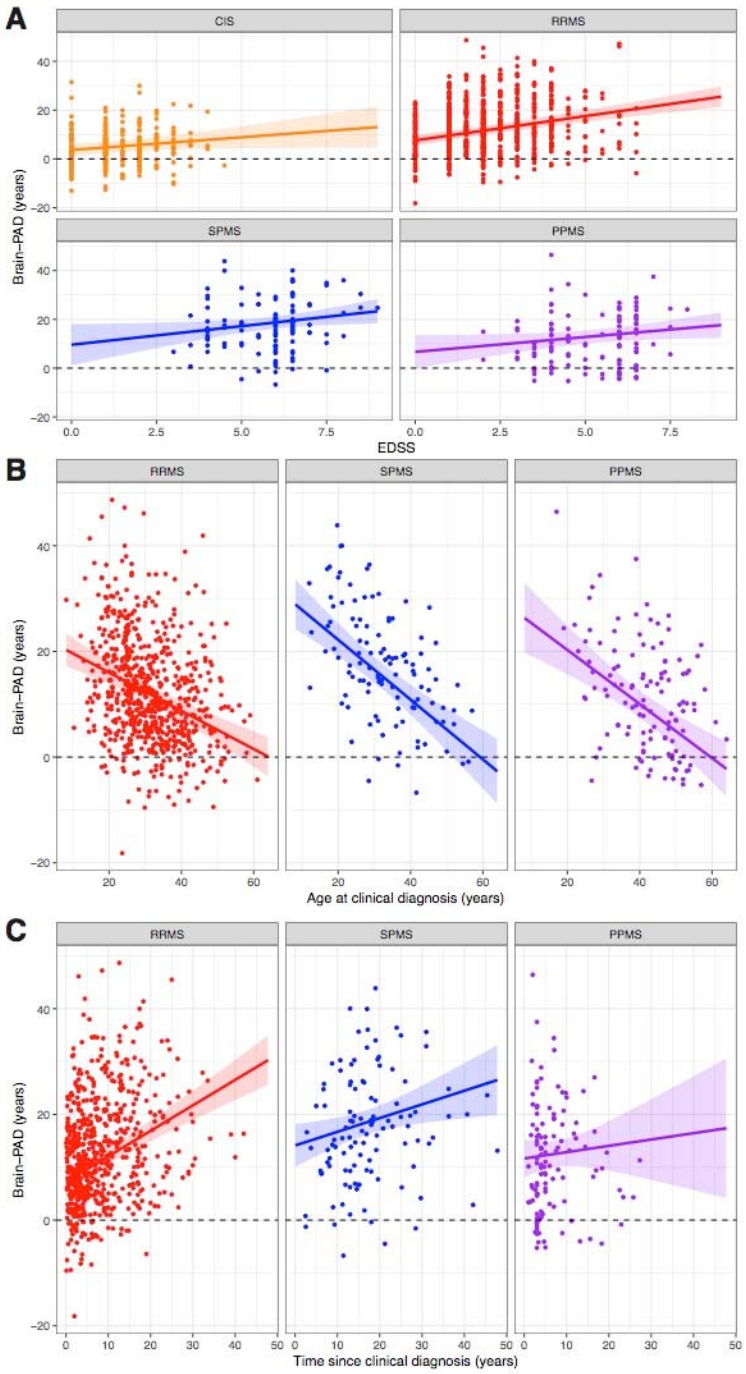


Figure 2. Scatterplot of brain-predicted age difference by age at diagnosis, time since diagnosis and EDSS score.

A) Baseline EDSS score (x-axis) and concurrent brain-PAD (y-axis). B) Age at clinical diagnosis at first scan (x-axis) and concurrent brain-PAD (y-axis). C) Time since diagnosis at baseline (x-axis) and concurrent brain-PAD (y-axis). Panels show patients with RRMS, SPMS and PPMS separately. Panels show patients with RRMS, SPMS and PPMS separately. Panels show patients with CIS, RRMS, SPMS and PPMS separately. Lines represented the linear regression lines calculated per group, and shaded areas are the 95% confidence intervals. CIS = clinically isolated syndrome,

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

RRMS = relapsing remitting MS, SPMS = secondary progressive MS, PPMS = primary progressive MS.

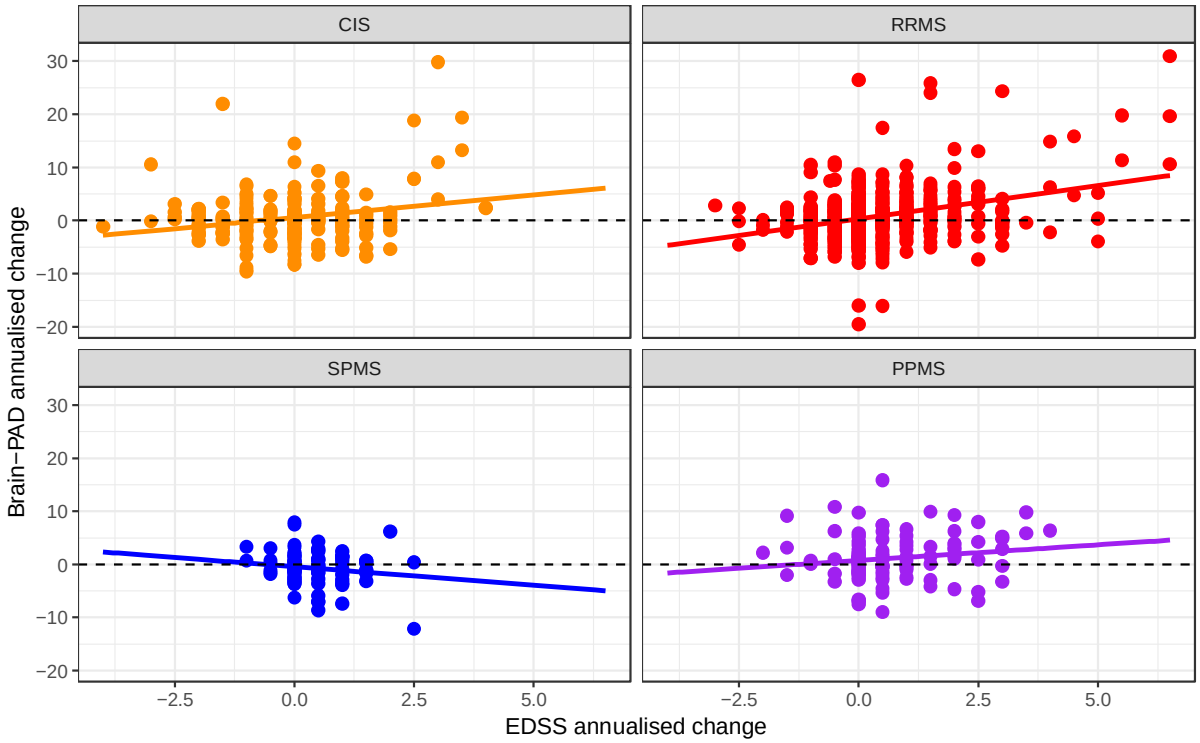


Figure 3. Scatterplot of annualised changed in EDSS score and brain-predicted age difference.

Panels show patients with CIS, RRMS, SPMS and PPMS separately, with annualised change in EDSS score between baseline and final follow-up (x-axis) and annualised change in brain-PAD between baseline and final follow-up (y-axis). Lines represented the linear regression lines calculated per group, and shaded areas are the 95% confidence intervals. CIS = clinically isolated syndrome, RRMS = relapsing remitting MS, SPMS = secondary progressive MS, PPMS = primary progressive MS.

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

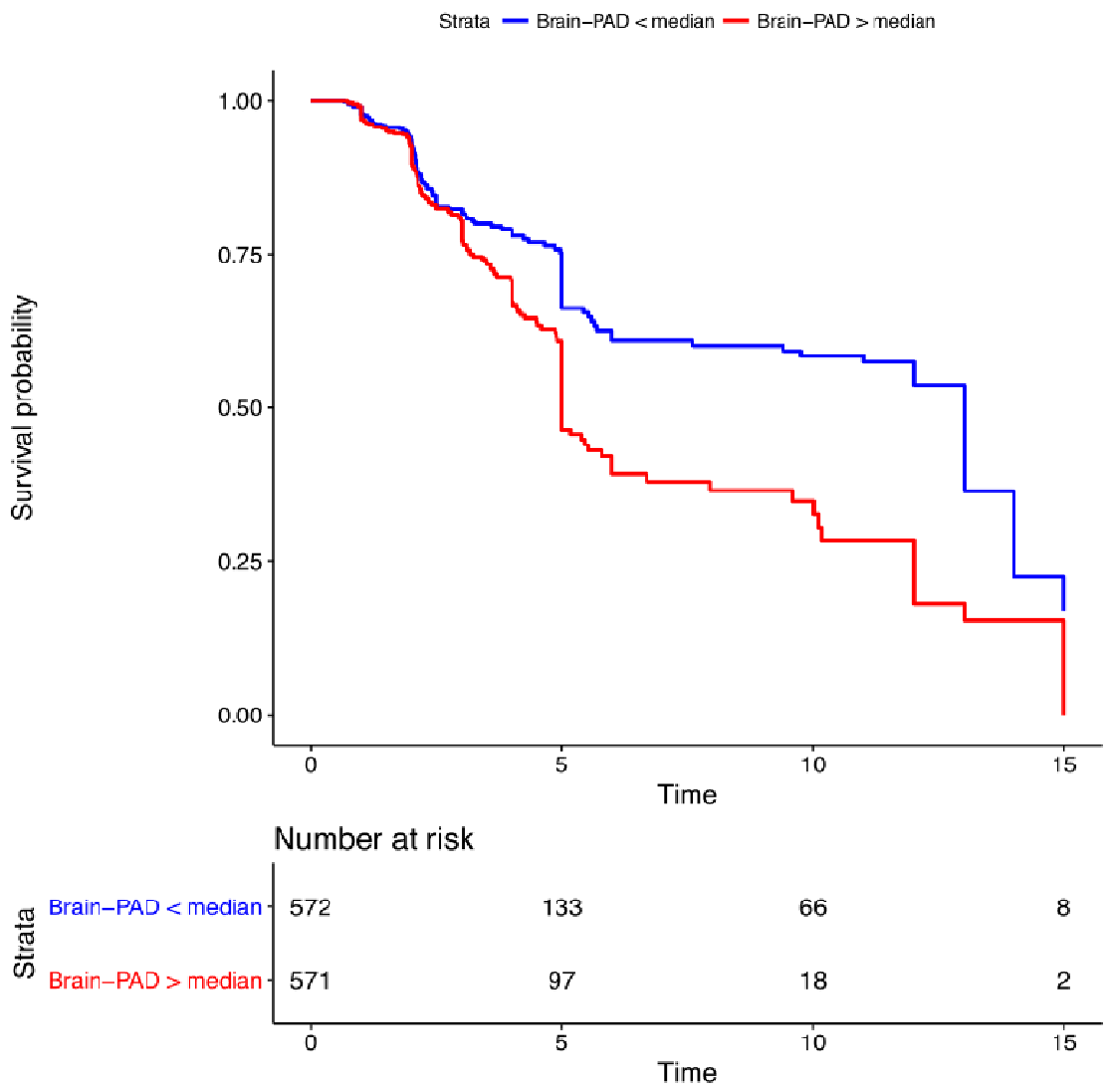


Figure 4. Time-to-EDSS progression survival curves based on baseline brain-PAD

Kaplan-Meier plot illustrating the relationship between brain-PAD at first scan and survival prior to an EDSS progression "event". Based on a median split of brain-PAD within MS/CIS patients (median brain-PAD = +9·68 years).

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

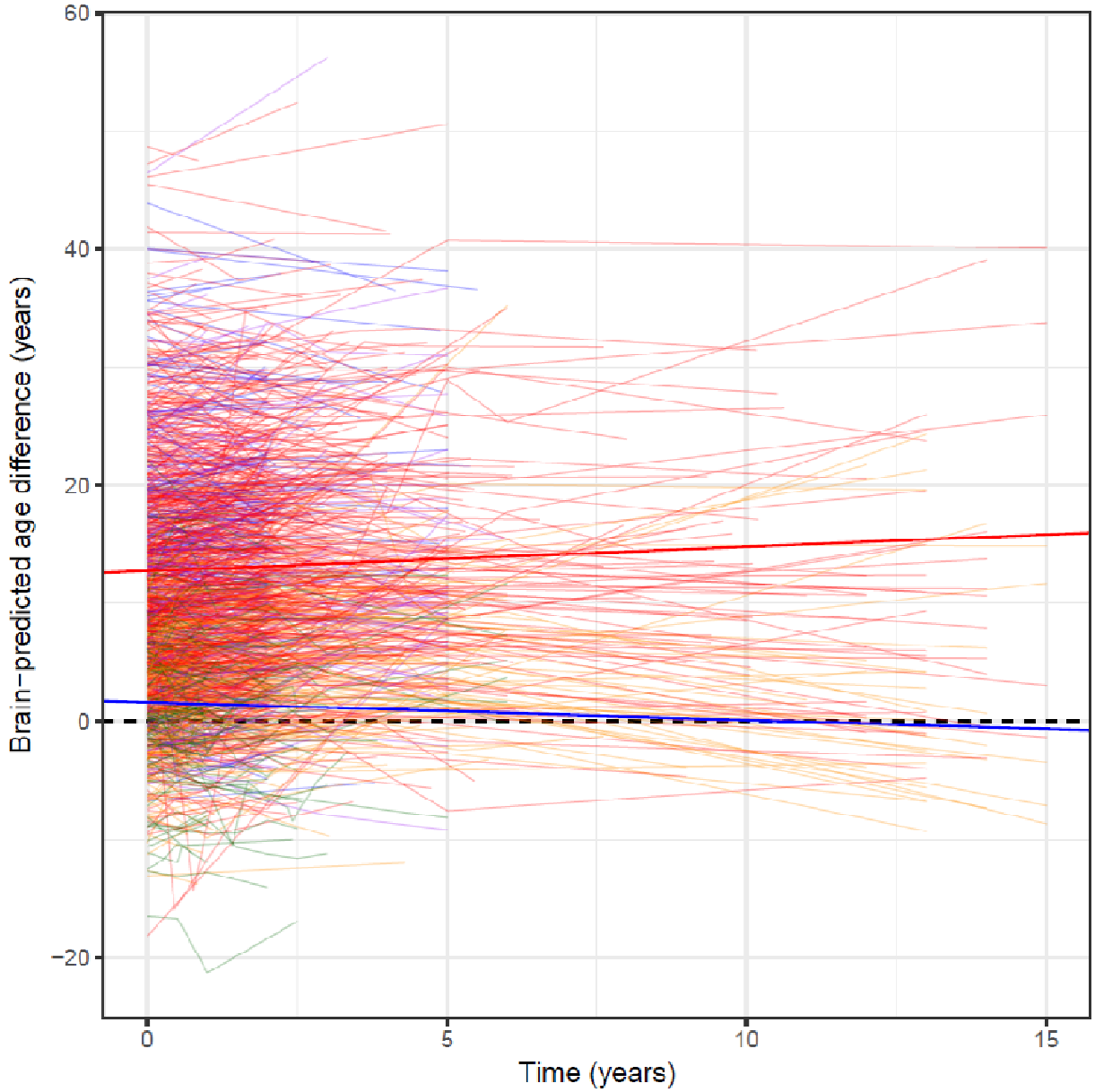


Figure 5. Individual trajectories of brain-predicted age difference by time from baseline. Lines show individual trajectories of brain-PAD scores over the longitudinal study period, coloured according to group (HC = green, CIS = orange, RRMS = red, SPMS = blue, PPMS = purple) Time (from baseline scan) in years (x-axis) and brain-PAD (y-axis). The solid lines represent the average longitudinal slopes for HCs (blue) and all MS/CIS patients (red). The dashed line shows brain-PAD = 0 years.

Appendix

Table S1. MRI acquisition protocols

Centre	**London**					**Milan**		**Graz**	**Barcelona**		**Amsterdam**	**Rome**	**Siena**
Magnetic field	1.5 Tesla			3 Tesla		1.5 Tesla	3 Tesla	3 Tesla	1.5 Tesla	3 Tesla	1.5 Tesla	1.5 Tesla	1.5 Tesla
Vendor	GE	GE	GE	Philips	Siemens	Siemens	Philips	Siemens	Siemens	Siemens	Siemens	Siemens	Philips
Model	Signa	Signa	Signa	Achieva	Verio	Avanto	Intera	Tim Trio	Symphony	Tim Trio	Vision	Avanto	Gyroscan
Years of recruitment	1999-2006	1998-2005	1998-2010	2011-2016	2014-2016	2007-2014	2008-2015	2006-2013	2013-2016	2010-2016	2004-2007	2012-2016	1999-2013
Included studies	1	1	2	2	1	1	1	1	1	1	1	1	1
Acquisition dimensions	3D	3D	3D	3D	3D	3D	3D	3D	3D	3D	3D	3D	2D
Voxel dimensions (mm)	1.2x1.2x1.5	1.2x1.2x1.2	1.2x1.2x1.5	1x1x1	1x1x1	1x1x1	0.89x0.89x1	1x1x1	1x1x1	1x1x1.2	1x1x1	1x1x1	0.97x0.97x3
TR (ms)	13.3	14.3	29	6.8	2300	2000	25	1900	1980	2300	4000	9000	35
TE (ms)	4.2	5.1	15	3.1	2.98	3.93	4.6	2.6	3.1	2.98	20	89	10
Matrix size	256x256	256x256	256x256	256x256	256x256	256x224	256x256	176x221	256x256	256x240	180x256	192x256	256x256
Slices	124	156	124	256	160	208	220	256	176	128	256	160	50
	Numbers of participants												
Controls	44	0	0	98	8	0	0	0	0	0	0	0	0
CIS	24	4	70	1	0	11	0	95	0	84	4	1	2
RRMS	33	7	67	52	16	28	38	69	57	6	135	66	107
SPMS	0	0	0	32	1	8	16	9	5	0	36	4	0
PPMS	42	0	0	42	0	13	0	1	0	0	22	0	0
QC exclusions	0	0	0	0	0	0	0	0	0	0	13	0	0

CIS = clinically isolated syndrome; HC = healthy controls; ms = milliseconds; mm = millimetre; PPMS = primary-progressive multiple sclerosis; QC = quality control; RRMS = relapsing-remitting multiple sclerosis; SPMS = secondary-progressive multiple sclerosis; TE = echo time; TR = repetition time.

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S2. Data sources for healthy brain age training sample

Cohort	N	Age mean (SD)	Age range	Sex male/female	Repository details	Scanner (Field strength)	Scan	Voxel dimensions
ABIDE (Autism Brain Imaging Data Exchange)	184	25.93 (6.66)	18-48	161/23	INDI	Various (all 3T)	MPRAGE	Various
Beijing Normal University	179	21.25 (1.92)	18-28	72/107	INDI	Siemens (3T)	MPRAGE	1.33 x 1.0 x 1.0
Berlin School of Brain & Mind	49	30.99 (7.08)	20-60	24/25	INDI	Siemens Tim Trio (3T)	MPRAGE	1.0 x 1.0 x 1.0
Computer Aided Diagnosis of Dementia	12	62.33 (6.26)	55-79	9/3	http://caddementia.grand-challenge.org	GE Signa (3T)	3D IR-FSPGR	0.9 x 0.9 x 1.0
Cleveland Clinic	31	43.55 (11.14)	24-60	11/20	INDI	Siemens Tim Trio (3T)	MPRAGE	2.0 x 1.0 x 1.2
ICBM (International Consortium for Brain Mapping)	322	24.84 (5.14)	24-60	177/145	LONI IDA	Siemens Magnetom (1.5T)	MPRAGE	1.0 x 1.0 x 1.0
IXI (Information eXtraction from Images)	561	48.62 (16.49)	20-86	250/311	http://brain-development.org/ixi-dataset/	Philips Intera (3T); Philips Gyroscan Intera (1.5T); GE Signa (1.5T)	T1-FFE; MPRAGE	0.9375 x 0.9375 x 1.2
MCIC (MIND Clinical Imaging Consortium)	93	32.49 (11.95)	18-60	64/29	COINS	Siemens Sonata/Trio (1.5/3T); GE Signa (1.5T)	MPRAGE; SPGR	0.625 x 0.625 x 1.5
MIRIAD (Minimal Interval Resonance Imaging in Alzheimer's Disease)	23	69.66 (7.18)	58-85	12/11	https://www.ucl.ac.uk/drc/research/methods/minimal-interval-resonance-imaging-alzheimers-disease-miriad	GE Signa (1.5T)	3D IR-FSPGR	0.9375 x 0.93751 x1.5
NEO2012 (Adelstein, 2011)	39	29.59 (8.38)	20-49	18/21	INDI	Siemens Allegra (3T)	MPRAGE	1.0 x 1.0 x 1.0
Nathan Kline Institute (NKI) / Rockland	160	41.49 (18.08)	18-85	96/64	INDI	Siemens Tim Trio (3T)	MPRAGE	1.0 x 1.0 x 1.0
OASIS (Open Access Series of Imaging Studies)	288	44.06 (23.04)	18-90	106/188	http://www.oasis-brains.org/	Siemens Vision (1.5T)*	MPRAGE	1.0 x 1.0 x 1.25
WUSL (Power, 2012)	24	23.04 (1.42)	20-24	4/20	INDI	Siemens Tim Trio (3T)	MPRAGE	1.0 x 1.0 x 1.0
TRAIN-39	36	22.67 (2.56)	18-28	11/25	INDI	Siemens Allegra (3T)	MPRAGE	1.33 x 1.33 x 1.3
Training set total	2001	36.95 (18.12)	18-90	1016/985	-	-	-	-

INDI = International Neuroimaging Data-sharing Initiative (http://fcon_1000.projects.nitrc.org)
COINS = Collaborative Informatics and Neuroimaging Suite (http://coins.mrn.org)
LONI = Laboratory of Neuro Imaging Image & Data Archive (https://ida.loni.usc.edu/)
ABIDE consortiums comprising data from various sites with different scanners/parameters
***OASIS scans were acquired four times and then averaged to increase signal-to-noise ratio.**

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S3. Post-hoc comparison of differences in brain-PAD between subtypes

Group comparison	Mean difference (brain-PAD years)	95% Confidence intervals	P-value
CIS-HC	3.46	0.93, 5.98	0.0018
RRMS-HC	11.38	9.11, 13.65	<0.0001
SPMS-HC	16.12	12.96, 19.27	<0.0001
PPMS-HC	10.09	7.00, 13.17	<0.0001
RRMS-CIS	7.92	6.17, 9.68	<0.0001
SPMS-CIS	12.66	9.85, 15.46	<0.0001
PPMS-CIS	6.63	3.90, 9.36	<0.0001
SPMS-RRMS	4.74	2.16, 7.32	<0.0001
PPMS-RRMS	-1.29	-3.79, 1.20	0.62
PPMS-SPMS	-6.03	-9.35, -2.71	<0.0001

CIS = clinically isolated syndrome; HC = healthy controls; PPMS = primary-progressive multiple sclerosis; RRMS = relapsing-remitting multiple sclerosis; SPMS = secondary-progressive multiple sclerosis.

Table S4. Variance explained in brain-predicted age difference by different predictors from hierarchical partitioning of variance.

Predictor variable	Unique R^2	Percentage of model variance explained
Chronological age	0.090	18.4%
Sex	0.005	1.1%
GM volume/ICV	0.152	31.2%
WM volume/ICV	0.046	9.4%
CSF volume/ICV	0.195	39.9%
Total	0.488	100%

CSF = cerebrospinal fluid; GM = grey matter; ICV = Intracranial volume; WM = white matter.

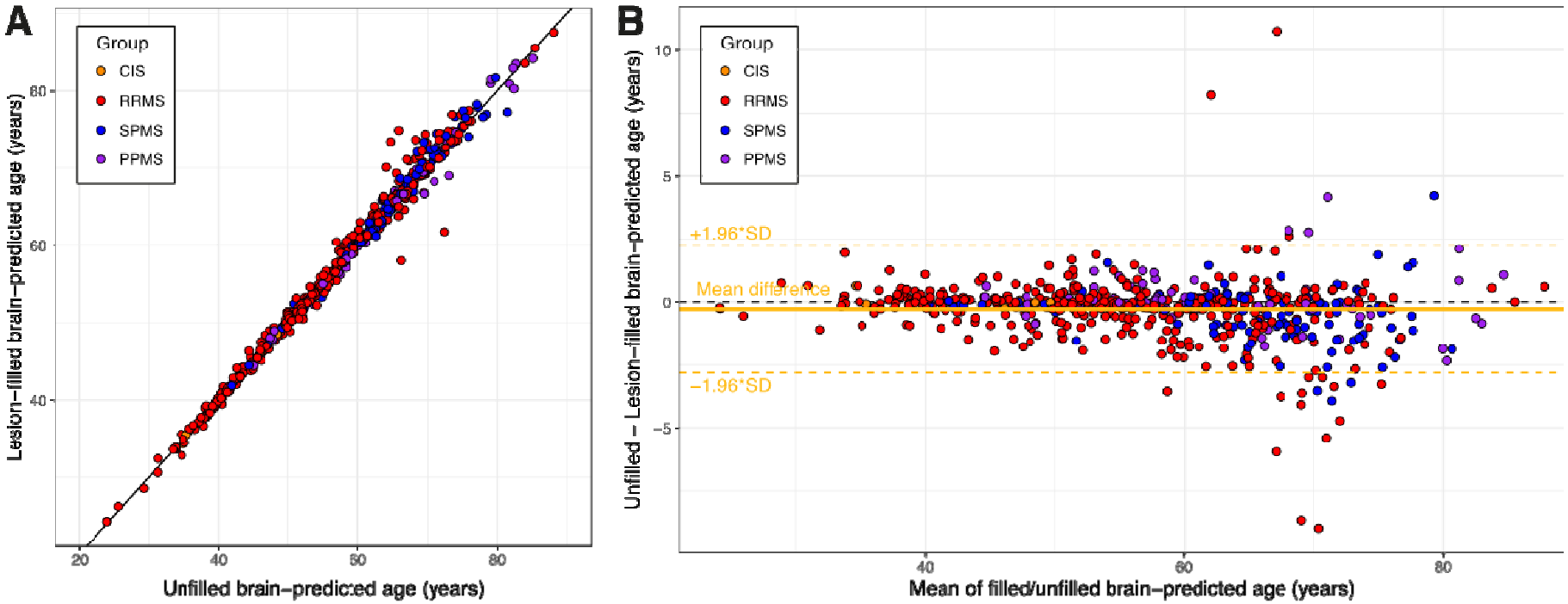


Figure S1. Impact of lesion filling procedure on brain-PAD.

A) Scatterplot showing brain-predicted age derived from original, 'unfilled' T1-weighted MRI scans (x-axis), plotted against brain-predicted age generated from T1-weighted MRIs that had undergone the automated lesion 'filling' procedure. Points are coloured according to MS subtype or CIS. B Bland-Altman plot of brain-predicted age from unfilled T1-weighted MRI scans and brain-predicted ages generated from 'filled' T1-weighted MRIs. The plot shows the mean value from the two measures for each participant (x-axis) and the difference between the two measures (y-axis). The black dashed line is the line of equality. The mean difference line is plotted in darkgoldenrod1 (mean difference = -0.28 years), along with the limits of agreement (±1.96 * standard deviation of differences, dashed). Total n = 575, with CIS = 8, RRMS = 382, SPMS = 119, PPMS = 66.

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MAGNIMS Study Group: Steering Committee Members

Alex Rovira (co-chair): MR Unit and Section of Neuroradiology, Department of Radiology, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain

Christian Enzinger (co-chair): Department of Neurology, Medical University of Graz, Graz, Austria

Frederik Barkhof: Queen Square Multiple Sclerosis Centre, UCL Institute of Neurology, University College London, London, UK

Olga Ciccarelli: Queen Square Multiple Sclerosis Centre, UCL Institute of Neurology, University College London, London, UK

Massimo Filippi: Neuroimaging Research Unit, Institute of Experimental Neurology, Division of Neuroscience, San Raffaele Scientific Institute, Vita-Salute San Raffaele University, Milan, Italy

Nicola De Stefano: Department of Medicine, Surgery and Neuroscience, University of Siena, Siena, Italy

Ludwig Kappos: Department of Neurology, University Hospital, Kantonsspital, Basel, Switzerland

Jette Frederiksen: The MS Clinic, Department of Neurology, University of Copenhagen, Glostrup Hospital, Denmark

Jaqueline Palace: Centre for Functional Magnetic Resonance Imaging of the Brain, University of Oxford, UK

Maria A Rocca: Neuroimaging Research Unit, Institute of Experimental Neurology, Division of Neuroscience, San Raffaele Scientific Institute, Vita-Salute San Raffaele University, Milan, Italy

Jaume Sastre-Garriga: Department of Neurology/Neuroimmunology, Multiple Sclerosis Centre of Catalonia (CEMCAT), Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain

Hugo Vrenken: Department of Radiology and Nuclear Medicine, MS Center Amsterdam, Amsterdam, The Netherlands

Tarek Yousry: NMR Research Unit, Institute of Neurology, University College London, London, UK

Claudio Gasperini: Department of Neurology and Psychiatry, University of Rome Sapienza, Rome, Italy.

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

R Notebook used for statistical analysis

James Cole - March 2019. Built with R version 3.5.2

This is Notebook contains the final brain age analysis of MS patient data and controls from the UCL cohort, the MAGNIMS consortium and the Imperial College London PET study (n=25). The analysis uses brain-predicted age difference (brain-PAD) to look at brain ageing in the context of MS. The brain-PAD values were generated in PRONTO, using an independent healthy (n=2001) training dataset, and the values were corrected for proportional bias using the intercept and slope of the age by brain-predicted age regression in the training dataset.

Initial set up of analysis

Clear workspace, load libraries, set colour palette

```
rm(list = ls()) ## clear workspace
library(car)
library(data.table)
library(dplyr)
library(emmeans)
library(ggstance)
library(ggplot2)
library(grDevices)
library(gridExtra)
library(hier.part)
library(jtools)
library(knitr)
library(lmerTest)
library(MASS)
library(metafor)
library(plotrix)
library(plyr)
library(pryr)
library(psych)
library(qwraps2)
library(reshape)
library(reshape2)
library(RColorBrewer)
library(scales)
library(stringr)
library(survival)
library(survminer)
library(tidyr)
ms.palette <- c("darkgreen", "darkorange", "red", "blue", "purple") # define MS
colour scheme for groups
sessionInfo()
```

```
## R version 3.5.2 (2018-12-20)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS High Sierra 10.13.4
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
## locale:
## [1] en_GB.UTF-8/en_GB.UTF-8/en_GB.UTF-8/C/en_GB.UTF-8/en_GB.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] tidyr_0.8.2        survminer_0.4.3    ggpubr_0.2
##  [4] magrittr_1.5       survival_2.43-3    stringr_1.3.1
##  [7] scales_1.0.0       RColorBrewer_1.1-2 reshape2_1.4.3
## [10] reshape_0.8.8      qwraps2_0.3.0      psych_1.8.10
## [13] pryr_0.1.4         plyr_1.8.4         plotrix_3.7-4
## [16] metafor_2.0-0      MASS_7.3-51.1      lmerTest_3.0-1
## [19] lme4_1.1-19        Matrix_1.2-15      knitr_1.21
## [22] jtools_1.1.1       hier.part_1.0-4    gtools_3.8.1
## [25] gridExtra_2.3      ggplot2_3.1.0      ggstance_0.3.1
## [28] emmeans_1.3.1      dplyr_0.7.8        data.table_1.11.8
## [31] car_3.0-2          carData_3.0-2
##
## loaded via a namespace (and not attached):
##  [1] splines_3.5.2     assertthat_0.2.0  cellranger_1.1.0
##  [4] yaml_2.2.0        backports_1.1.3   numDeriv_2016.8-1
##  [7] pillar_1.3.1      lattice_0.20-38   glue_1.3.0
## [10] digest_0.6.18     minqa_1.2.4       colorspace_1.3-2
## [13] cmprsk_2.2-7      htmltools_0.3.6   pkgconfig_2.0.2
## [16] broom_0.5.1       haven_2.0.0       purrr_0.2.5
## [19] xtable_1.8-3      mvtnorm_1.0-8     km.ci_0.5-2
## [22] openxlsx_4.1.0    rio_0.5.16        KMsurv_0.1-5
## [25] tibble_1.4.2      generics_0.0.2    withr_2.1.2
## [28] lazyeval_0.2.1    cli_1.0.1         mnormt_1.5-5
## [31] crayon_1.3.4      readxl_1.2.0      estimability_1.3
## [34] evaluate_0.12     nlme_3.1-137      forcats_0.3.0
## [37] foreign_0.8-71    tools_3.5.2       hms_0.4.2
## [40] munsell_0.5.0     zip_1.0.0         bindrcpp_0.2.2
## [43] compiler_3.5.2    rlang_0.3.0.1     grid_3.5.2
## [46] nloptr_1.2.1      rmarkdown_1.11    gtable_0.2.0
## [49] codetools_0.2-16  abind_1.4-5       curl_3.2
## [52] R6_2.3.0          zoo_1.8-4         survMisc_0.5.5
## [55] bindr_0.1.1       stringi_1.2.4     parallel_3.5.2
## [58] Rcpp_1.0.0        tidyselect_0.2.5  xfun_0.4
```

Get data from CSV and define longitudinal data.frames

```
setwd('/Users/jcole/Work/Brain ageing/Collaborations/MS')
df <- read.csv("MS_brain_age_final_long.csv")
df$subtype <- factor(df$subtype, levels = c("control", "CIS", "RRMS", "SPMS",
"PPMS")) # reorder subtype factor to put controls first
df$Cohort <- recode(df$Cohort, JR1 = "Imperial", C0 = "UCL0", C1 = "UCL1", C2 =
"UCL2", C3 = "UCL3", C4 = "UCL4" ,C5 = "UCL5", C6 = "UCL6", C7 = "UCL7", A =
"Amsterdam", B = "Barcelona", G = "Graz", M = "Milan", R = "Rome", S = "Siena")
```

Exclude participants with errors in the database & correct time since diagnosis errors

```
tmp <- df[df$Cohort == 'Amsterdam',] %>% group_by(PatientID) %>%
dplyr::summarize(sd = sd(age_at_scan)) %>% arrange(desc(sd)) %>% filter(sd > 2)
excluded_IDs <- sort(tmp$PatientID)
# excluded_IDs <- unique(df[which(df$age_at_scan < df$age_at_baseline_scan1),1])
excluded_scans <- (df %>% filter(str_detect(PatientID, paste(excluded_IDs,
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
collapse = "|"))))["ScanName"]
df <- df %>%
  filter(!str_detect(PatientID, paste(excluded_IDs, collapse = "|")))
rm(tmp)
```

There were 13 subjects with 38 scans excluded in total. Data entry errors in original spreadsheet; age at baseline not consistent within subject.

Load data for lesion filling analysis

```
## read in data
lesion_df <- read.csv(file = '~/Work/Brain
ageing/Collaborations/MS/MAGNIMS/MAGNIMS20171224_final.csv')
lesion_df$subtype <- factor(lesion_df$subtype, levels = c("control", "CIS",
"RRMS", "SPMS", "PPMS"))
## create ICV ratio variables
lesion_df <- lesion_df %>%
  mutate(ICV = GMV + WMV + CSFV) %>%
  mutate(gm_icv_ratio = GMV/ICV) %>%
  mutate(wm_icv_ratio = WMV/ICV)
```

Generate baseline only data.frame and show data frame structure

```
df.bl <- df[(df$age_at_baseline_scan1 == df$age_at_scan),] # define baseline
data.frame
str(df)

## 'data.frame':    3565 obs. of  32 variables:
##  $ PatientID                    : Factor w/ 1367 levels "AMSTERDAM_4001",..: 1 1 1 2 2 2 3 3 3 4 ...
##  $ Cohort                       : Factor w/ 15 levels "Amsterdam","Barcelona",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ scanner_status               : Factor w/ 4 levels "1.5T","1.5T_post_2004",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ gender                       : Factor w/ 2 levels "female","male": 2 2 2 1 1 1 1 1 1 1 ...
##  $ control0ppms1cisoforever2other3   : int  3 3 3 3 3 3 3 3 3 3 ...
##  $ control0rest1                : Factor w/ 2 levels "control","MS": 2 2 2 2 2 2 2 2 2 2 ...
##  $ control0ms1cis2              : Factor w/ 3 levels "CIS","control",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ age_at_baseline_scan1        : num  57.8 57.8 57.8 44.4 44.4 ...
##  $ disease_duration_at_baseline_scan1: num  6.8 6.8 6.8 6.38 6.38 ...
##  $ DMT_YesNoNA                  : Factor w/ 2 levels "NO","YES": 1 1 1 1 1 1 2 2 2 2 ...
##  $ DMTYes1                      : Factor w/ 2 levels "No treatment",..: 1 1 1 1 1 1 2 2 2 2 ...
##  $ EDSSbaseline                 : num  2.5 2.5 2.5 3 3 3 4 4 4 2.5 ...
##  $ NoScans                      : int  3 3 3 3 3 3 3 3 3 2 ...
##  $ FUTimeMax                    : num  2.32 2.32 2.32 2.35 2.35 ...
##  $ ScanName                     : Factor w/ 3603 levels "AMSTERDAM_4001_baseline_t1",..: 1 2 3 4 5 6 7 8
9 10 ...
##  $ age_at_scan                  : num  57.8 59.1 60.1 44.4 45.5 ...
##  $ EDSSatScan                   : num  2.5 3.5 3 3 2 2.5 4 2.5 4 2.5 ...
##  $ BrainAge                     : num  69.6 73.4 71 58.1 61.8 ...
##  $ BrainPAD                     : num  11.8 14.3 10.8 13.7 16.3 ...
##  $ subtype                      : Factor w/ 5 levels "control","CIS",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ disease_onset_age            : num  51 51 51 38 38 ...
##  $ disease_duration             : num  6.8 6.8 6.8 6.38 6.38 ...
##  $ interval                     : num  0 1.3 2.32 0 1.16 ...
##  $ scan_number                  : Factor w/ 10 levels "scan1","scan10",..: 1 3 4 1 3 4 1 3 4 1 ...
##  $ GM_vol                       : num  0.574 0.563 0.573 0.622 0.605 0.59 0.647 0.643 0.633 0.648 ...
##  $ WM_vol                       : num  0.536 0.532 0.52 0.43 0.423 0.413 0.404 0.399 0.387 0.599 ...
##  $ CSF_vol                      : num  0.352 0.349 0.367 0.293 0.309 0.333 0.355 0.358 0.386 0.271 ...
##  $ WBV                          : num  1.11 1.09 1.09 1.05 1.03 ...
##  $ ICV                          : num  1.46 1.44 1.46 1.34 1.34 ...
##  $ gm_vol_ratio_icv             : num  0.393 0.39 0.392 0.462 0.453 ...
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
##  $ wm_vol_ratio_icv          : num  0.367 0.368 0.356 0.32 0.316 ...
##  $ csf_vol_ratio_icv         : num  0.241 0.242 0.251 0.218 0.231 ...
```

Basic stats

Total number of subjects, and by group

The total number of subjects included was n = 1354

The total number of MS patients (including CIS) was n = 1204 and healthy controls was n = 150

Number of scans in total and per group

Total number of scans = 3565

Number of people with 2 or more scans

```
df.bl %>%
  filter(NoScans >= 2) %>%
  group_by(control0rest1) %>%
  tally()

## # A tibble: 2 x 2
##   control0rest1     n
##   <fct>       <int>
## 1 control       111
## 2 MS           1155
```

Number of people with 3 or more scans

```
df.bl %>%
  filter(NoScans >= 3) %>%
  group_by(control0rest1) %>%
  tally()

## # A tibble: 2 x 2
##   control0rest1     n
##   <fct>       <int>
## 1 control        64
## 2 MS            509
```

Generate demographics table using qwarps2

```
options(qwraps2_markup = "markdown", digits = 2)

table1 <-
  list("N" =
         list("Control" = ~ qwraps2::n_perc0(control0rest1 == "control", na_rm =
T),
              "MS"   = ~ qwraps2::n_perc0(control0rest1 == "MS", na_rm = T)),
     "Gender" =
       list("Female" = ~ qwraps2::n_perc0(gender == "female", show_symbol = T),
            "Male"   = ~ qwraps2::n_perc0(gender == "male", show_symbol = T)),
        "Number of scans" =
       list("min" = ~ min(NoScans),
            "max" = ~ max(NoScans),
            "mean (sd)" = ~ qwraps2::mean_sd(NoScans)),
       "Age at baseline scan (years)" =
       list("min" = ~ min(age_at_baseline_scan1),
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
            "max" = ~ max(age_at_baseline_scan1),
            "mean (sd)" = ~ qwraps2::mean_sd(age_at_baseline_scan1)),
       "Brain-predicted age at baseline scan (years)" =
       list("min" = ~ min(BrainAge),
            "max" = ~ max(BrainAge),
            "mean (sd)" = ~ qwraps2::mean_sd(BrainAge)),
       "Disease duration at baseline (years)" =
       list("min" = ~ min(disease_duration, na.rm = T),
            "max" = ~ max(disease_duration, na.rm = T),
            "mean (sd)" = ~ qwraps2::mean_sd(disease_duration, na_rm = T, show_n
= "never")),
        "EDSS at baseline " =
       list("min" = ~ min(EDSSbaseline, na.rm = T),
            "max" = ~ max(EDSSbaseline, na.rm = T),
            "mean (sd)" = ~ qwraps2::mean_sd(EDSSbaseline, na_rm = T, show_n =
"never")),
       "MS subtype" =
       list("CIS" = ~ qwraps2::n_perc0(subtype == "CIS", show_symbol = T),
            "RRMS"  = ~ qwraps2::n_perc0(subtype == "RRMS", show_symbol = T),
            "SPMS"  = ~ qwraps2::n_perc0(subtype == "SPMS", show_symbol = T),
            "PPMS"  = ~ qwraps2::n_perc0(subtype == "PPMS", show_symbol = T)),
       "Treatment" =
         list("Yes" = ~ qwraps2::n_perc0(DMT_YesNoNA == "YES", na_rm = T,
show_symbol = T),
              "No"  = ~ qwraps2::n_perc0(DMT_YesNoNA == "NO", na_rm = T,
show_symbol = T))
       )

print(summary_table(dplyr::group_by(df.bl, control0rest1), table1),
      rtitle = "Sample Chararcteristics",
      cnames = c("Controls", "MS patients"))

##
##
## |Sample Chararcteristics                    |Controls          |MS patients        |
## |:-----------------------------------------|:-------------------|:-------------------|
## |**N**                                |         |         |
## |   Control                      |150 (100)         |0 (0)             |
## |   MS                           |0 (0)             |1,204 (100)        |
## |**Gender**                           |         |         |
## |   Female                       |82 (55%)          |771 (64%)          |
## |   Male                         |68 (45%)          |433 (36%)          |
## |**Number of scans**                      |         |         |
## |   min                          |1                 |1                  |
## |   max                          |10                |7                  |
## |   mean (sd)                    |2.82 &plusmn; 1.90  |2.61 &plusmn; 1.01  |
## |**Age at baseline scan (years)**         |         |         |
## |   min                          |23                |15                 |
## |   max                          |66                |68                 |
## |   mean (sd)                    |37.29 &plusmn; 9.96  |39.41 &plusmn; 10.76 |
## |**Brain-predicted age at baseline scan (years)** |         |         |
## |   min                          |14.5              |7.4                |
## |   max                          |70                |92                 |
## |   mean (sd)                    |38.43 &plusmn; 11.12 |50.27 &plusmn; 14.90 |
## |**Disease duration at baseline (years)**     |         |         |
## |   min                          |Inf               |0                  |
## |   max                          |-Inf              |48                 |
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
## |   mean (sd)                    |NaN &plusmn;  NA     |7.26 &plusmn; 7.96   |
## |**EDSS at baseline **                     |           |           |
## |   min                          |Inf                  |0                    |
## |   max                          |-Inf                 |9                    |
## |   mean (sd)                    |NaN &plusmn;  NA     |2.60 &plusmn; 1.95   |
## |**MS subtype**                            |           |           |
## |   CIS                          |0 (0%)               |296 (25%)            |
## |   RRMS                         |0 (0%)               |677 (56%)            |
## |   SPMS                         |0 (0%)               |111 (9%)             |
## |   PPMS                         |0 (0%)               |120 (10%)            |
## |**Treatment**                             |           |           |
## |   Yes                          |0 (NaN%)             |475 (41%)            |
## |   No                           |0 (NaN%)             |675 (59%)            |
```

Need to get treatment NAs using table()

```
table(df.bl$DMT_YesNoNA, df.bl$control0rest1, useNA = "ifany")[,"MS"]

##   NO  YES <NA>
##  675  475   54
```

Need to get length of follow-up from longitudinal database

Length of follow-up

```
options(digits = 3) ## return digits option to default
df %>%
  filter(NoScans >= 2) %>%
  group_by(PatientID) %>%
  slice(which.max(interval)) %>%
  # top_n(n = 1, wt = interval) %>%
  group_by(control0rest1) %>%
  dplyr::summarise(mean(interval), sd(interval), min(interval), max(interval))

## # A tibble: 2 x 5
##   control0rest1 `mean(interval)` `sd(interval)` `min(interval)`
##   <fct>                   <dbl>         <dbl>         <dbl>
## 1 control                  1.97          1.38          0.5
## 2 MS                       3.41          3.15          0.233
## # ... with 1 more variable: `max(interval)` <dbl>

options(digits = 7) ## return digits option to default
```

Baseline brain-age analysis

```
  describeBy(df.bl$BrainPAD, df.bl$control0rest1, mat = T, digits = 3) # brain-PAD by MS patient vs. controls

##     item  group1 vars    n   mean     sd median trimmed   mad     min
## X11    1 control    1  150  1.135  6.738  0.642   1.173 6.510 -16.507
## X12    2      MS    1 1204 10.875 10.237  9.443  10.246 9.684 -18.181
##       max  range  skew kurtosis    se
## X11 22.090 38.597 0.032    0.072 0.550
## X12 48.666 66.847 0.600    0.315 0.295
```

Estimated marginal means

Generate EMMs for all MS/CIS and healthy controls. LME adjusting for age, gender, ICV, cohort and scanner status.

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
fit <- (lmer(BrainPAD ~ control0rest1 + age_at_baseline_scan1 + gender + ICV +
(1|Cohort/scanner_status), data = df.bl))
emmeans(object = fit, specs = c("control0rest1", "age_at_baseline_scan1",
"ICV"))

##  control0rest1 age_at_baseline_scan1     ICV   emmean      SE    df
##  control                 39.17223 1.395055 -0.01176 1.485229 32.69
##  MS                      39.17223 1.395055 11.81573 1.230551 11.99
##   lower.CL  upper.CL
##  -3.034578  3.011058
##   9.134334 14.497117
##
## Results are averaged over the levels of: gender
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
```

Effects of cohort and scanner status on brain-PAD.

The significant interaction motivates nesting in the LME.

```
with(df.bl, anova(lm(BrainPAD ~ Cohort * scanner_status)))

## Analysis of Variance Table
##
## Response: BrainPAD
##                    Df Sum Sq Mean Sq  F value Pr(>F)
## Cohort             14  16901  1207.2  13.6521 <2e-16 ***
## scanner_status      3   1383   461.0   5.2129 0.0014 **
## Cohort:scanner_status  1   9149  9148.6 103.4597 <2e-16 ***
## Residuals        1335 118050    88.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Meta-analysis looking at all the separate cohorts with MS/CIS patients and controls

Check which cohorts contain healthy controls and patients.

```
table(df.bl$subtype, df.bl$Cohort)

##
##         Amsterdam Barcelona UCL0 UCL1 UCL2 UCL3 UCL4 UCL5 UCL6 UCL7 Graz
##   control       0       0    0    0    0   82   16   17   17   10    0
##   CIS           4      84    0   75    0    0    0    0    0   24   95
##   RRMS        131      63   21   77    0   28    0   33    0    0   69
##   SPMS         36       5    0    2    7   23    0    0    0    0    9
##   PPMS         22       0    0    0    9   14   19    0   42    0    1
##
##         Imperial Milan Rome Siena
##   control       8    0    0    0
##   CIS           0   11    1    2
##   RRMS         16   66   66  107
##   SPMS          1   24    4    0
##   PPMS          0   13    0    0
```

Create data.frame with summary data appropriate for meta-analysis.

```
tmp0 <- df.bl %>% group_by(Cohort, control0rest1) %>%
dplyr::summarise(length(control0rest1))
meta.cohorts <- as.list(tmp0[tmp0$control0rest1 == "control",1])$Cohort
tmp <- df.bl %>%
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
  filter(str_detect(Cohort, paste(meta.cohorts, collapse = "|")))
tmp1 <- tmp %>% group_by(Cohort, control0rest1) %>% dplyr::summarise(n = n(),
Mean = mean(BrainPAD), SD  =  sd(BrainPAD))
tmp2 <- dcast(tmp1, Cohort ~ control0rest1, value.var = "n")
tmp3 <- dcast(tmp1, Cohort ~ control0rest1, value.var = "Mean")
tmp4 <- dcast(tmp1, Cohort ~ control0rest1, value.var = "SD")
names(tmp2) <- c("Cohort", "control_n", "MS_n")
names(tmp3) <- c("Cohort", "control_mean", "MS_mean")
names(tmp4) <- c("Cohort", "control_sd", "MS_sd")
tmp1x <- tmp %>% group_by(Cohort) %>% dplyr::summarise(pooled.SD  =
sd(BrainPAD))
meta.df <- join_all(list(tmp2,tmp3,tmp4, tmp1x), by = 'Cohort', type = 'left')
rm(list = ls(pattern = 'tmp*')) # remove temporary data frames
```

Run meta-analysis using the metafor package, to fit a random-effects meta-analysis using REML.

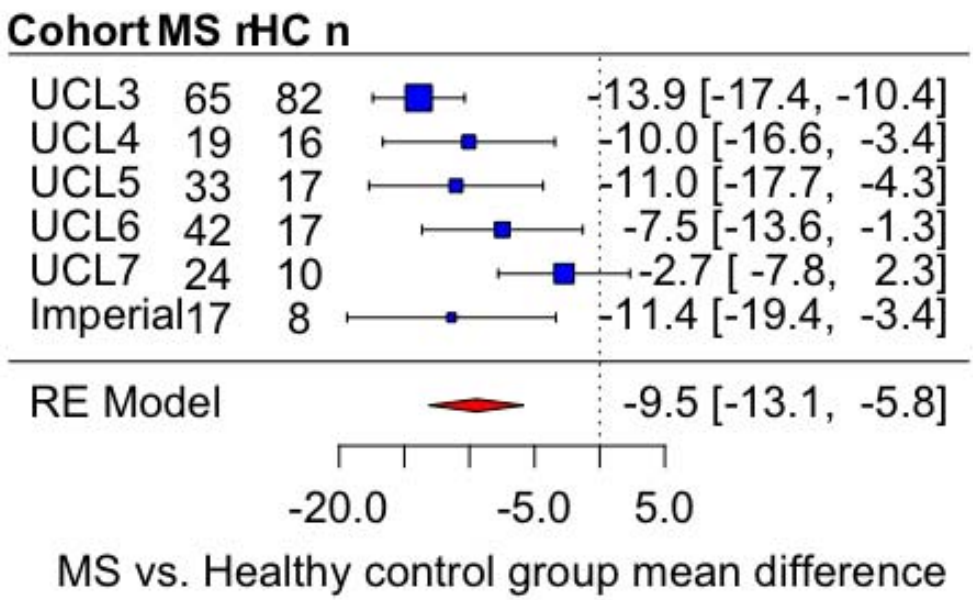
```
meta.df <- escalc(m1i = control_mean, sd1i = pooled.SD, n1i = control_n, m2i =
MS_mean, sd2i = pooled.SD, n2i = MS_n, measure = "MD", data = meta.df, digits =
2)
meta.results <- rma(yi, vi, data = meta.df, method = "REML")
print(meta.results)

##
## Random-Effects Model (k = 6; tau^2 estimator: REML)
##
## tau^2 (estimated amount of total heterogeneity): 11.8177 (SE = 13.0982)
## tau (square root of estimated tau^2 value):      3.4377
## I^2 (total heterogeneity / total variability):   59.03%
## H^2 (total variability / sampling variability):  2.44
##
## Test for Heterogeneity:
## Q(df = 5) = 13.4832, p-val = 0.0192
##
## Model Results:
##
## estimate      se     zval    pval    ci.lb   ci.ub
##  -9.4543  1.8649  -5.0695  <.0001  -13.1095  -5.7990  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

confint(meta.results)

##
##       estimate  ci.lb   ci.ub
## tau^2   11.8177 0.2530 84.1148
## tau      3.4377 0.5030  9.1714
## I^2(%)  59.0288 2.9923 91.1149
## H^2      2.4407 1.0308 11.2548
```

Forest plot of results

```
plot.forest %<a-% {
forest(meta.results, ilab = cbind(meta.df$MS_n, meta.df$control_n), ilab.xpos =
c(-30,-23), slab = meta.df$Cohort, digits = 1, xlab = "MS vs. Healthy control
group mean difference", steps = 6, col = "red", cex = 1.25, pch = 22, bg =
"blue"); text(c(-40, -30, -23), 7.6, c("Cohort", "MS n", "HC n"), font = 2, cex
= 1.25)
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
}
plot.forest
```




```
cairo_pdf("plots/forest_plot.pdf", 6,5)
plot.forest
dev.off()

## quartz_off_screen
##             2
```

Linear regression analysis restricted to cohort UCL3, includes covariates: age, gender, intracranial volume.

```
summary(lm(BrainPAD ~ control0rest1 + age_at_baseline_scan1 + gender + ICV, data
= subset(df.bl, df.bl$Cohort == "UCL3")))

##
## Call:
## lm(formula = BrainPAD ~ control0rest1 + age_at_baseline_scan1 +
##     gender + ICV, data = subset(df.bl, df.bl$Cohort == "UCL3"))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.9408  -5.1153   0.0774   4.7964  18.7904
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           3.44230    8.84583   0.389   0.6978
## control0rest1MS      15.53450    1.52984  10.154   <2e-16 ***
## age_at_baseline_scan1 -0.12140    0.06299  -1.927   0.0559 .
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
## gendermale          2.15828    1.86769   1.156   0.2498
## ICV                 0.75463    6.21246   0.121   0.9035
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.229 on 142 degrees of freedom
## Multiple R-squared:  0.4372, Adjusted R-squared:  0.4213
## F-statistic: 27.57 on 4 and 142 DF,  p-value: < 2.2e-16
```

Main result:

LME model to predict brain-PAD based on group

```
# LME model of control vs. MS, including Cohort and scanner status as random
effects
m1 <- lmer(BrainPAD ~ control0rest1 + age_at_baseline_scan1 + gender + ICV +
(1|Cohort/scanner_status), data = df.bl)
summary(m1)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## BrainPAD ~ control0rest1 + age_at_baseline_scan1 + gender + ICV +
##     (1 | Cohort/scanner_status)
##    Data: df.bl
##
## REML criterion at convergence: 9809.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.0598 -0.6584 -0.0873  0.5608  4.3393
##
## Random effects:
##  Groups              Name        Variance Std.Dev.
##  scanner_status:Cohort (Intercept) 20.3015  4.5057
##  Cohort              (Intercept)  0.6337  0.7961
##  Residual                         79.3950  8.9104
## Number of obs: 1354, groups:  scanner_status:Cohort, 19; Cohort, 15
##
## Fixed effects:
##                      Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)           0.40070    3.44160  761.33461   0.116 0.907343
## control0rest1MS      11.82749    0.99351 1322.26885  11.905  < 2e-16
## age_at_baseline_scan1  -0.09910    0.02558 1348.32958  -3.873 0.000112
## gendermale            1.65630    0.63972 1338.24863   2.589 0.009727
## ICV                   1.89331    2.21106 1334.85860   0.856 0.391991
##
## (Intercept)
## control0rest1MS      ***
## age_at_baseline_scan1 ***
## gendermale           **
## ICV
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) cn01MS ag___1 gndrml
## cntrl0rs1MS -0.250
## ag_t_bsln_1 -0.278 -0.163
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
## gendermale   0.475 -0.005  0.002
## ICV         -0.873  0.051  0.021 -0.606
```

```
round(confint(m1),2)
```

```
## Computing profile confidence intervals ...

##                    2.5 % 97.5 %
## .sig01              2.93   6.46
## .sig02              0.00   4.06
## .sigma              8.57   9.25
## (Intercept)        -6.33   7.12
## control0rest1MS     9.88  13.77
## age_at_baseline_scan1 -0.15  -0.05
## gendermale          0.41   2.91
## ICV                -2.44   6.22
```

Lesion filling

To establish whether using the FSL lesion filling software influences brain-predicted age values. This analysis was conducted only in UCL patients.

```
with(subset(lesion_df, lesion_df$subtype != "control"),
describeBy(filled_brain_age, subtype))
```

```
##
##  Descriptive statistics by group
## group: control
## NULL
## ---------------------------------------------------------
## group: CIS
##    vars n mean   sd median trimmed  mad  min   max range  skew kurtosis
## X1    1 8 42.4 6.18  42.57    42.4 7.87 34.3 49.69 15.39 -0.09    -1.98
##      se
## X1 2.18
## ---------------------------------------------------------
## group: RRMS
##    vars   n  mean    sd median trimmed   mad   min   max range  skew
## X1    1 382 54.78 11.46  54.79   54.78 12.12 24.26 87.53 63.27 -0.03
##    kurtosis   se
## X1    -0.44 0.59
## ---------------------------------------------------------
## group: SPMS
##    vars   n  mean   sd median trimmed  mad   min   max range  skew
## X1    1 119 64.62 9.24  65.96   65.41 8.45 40.33 81.64 41.31 -0.74
##    kurtosis   se
## X1    -0.03 0.85
## ---------------------------------------------------------
## group: PPMS
##    vars  n mean    sd median trimmed   mad  min   max range skew kurtosis
## X1    1 66 62.2 10.13  59.63   61.67 10.57 44.4 84.13 39.72 0.46    -0.61
##      se
## X1 1.25
```

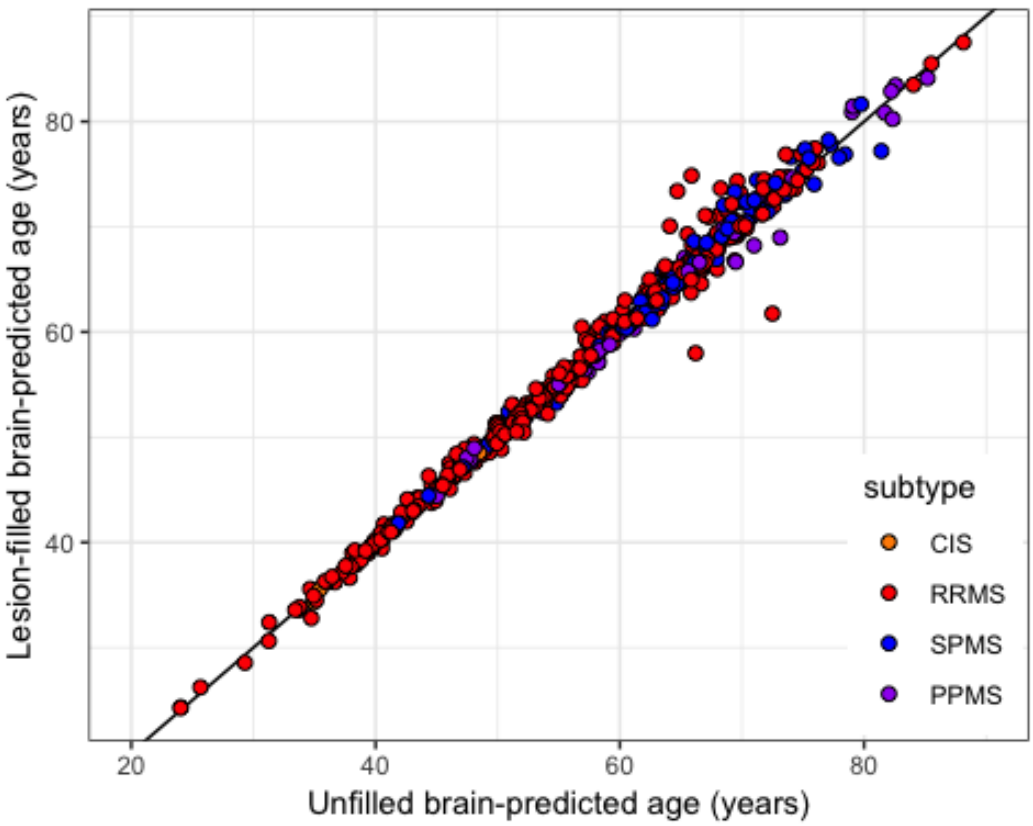
Correlation between brain-predicted age from filled and unfilled images: Pearson's r = 0.994. Median absolute error (MAE) between brain-predicted age from filled and unfilled images = 0.3717 years. Mean difference between brain-predicted age from filled and unfilled images = 0.28 years.

```
ggplot(data = subset(lesion_df, lesion_df$subtype != "control"), aes(x =
brain_age, y = filled_brain_age)) +
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
  geom_abline(slope = 1) +
  geom_point(pch = 21, aes(fill = subtype), size = 2) +
  labs(x = "Unfilled brain-predicted age (years)", y = "Lesion-filled brain-
predicted age (years)") +
  xlim(c(20,90)) +
  scale_fill_manual(values = ms.palette[-1]) +
  theme_bw() + theme(legend.position = c(0.9, 0.2))
```

Warning: Removed 1441 rows containing missing values (geom_point).




```
ggsave("~/Work/Brain
ageing/Collaborations/MS/plots/lesion_filling_brain_age_plot.pdf", width = 5,
height = 5, useDingbats = FALSE)
```

Warning: Removed 1441 rows containing missing values (geom_point).

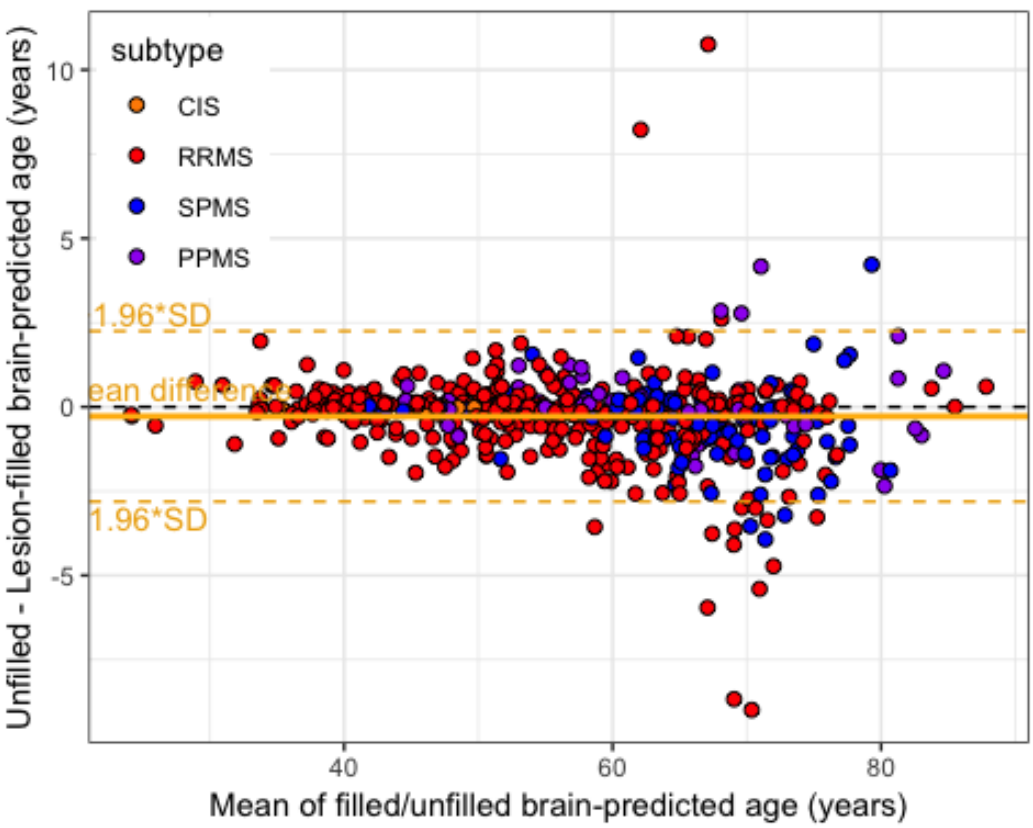
Bland-Altman plot

```
mean.diff <- mean(lesion_df$brain_age - lesion_df$filled_brain_age, na.rm = T)
sd.diff <- sd(lesion_df$brain_age - lesion_df$filled_brain_age, na.rm = T)
ggplot(data = subset(lesion_df, lesion_df$subtype != "control"), aes(x =
((brain_age + filled_brain_age)/2), y = brain_age - filled_brain_age)) +
  geom_abline(slope = 0, lty = 2) +
  geom_point(pch = 21, aes(fill = subtype), size = 2) +
  geom_hline(yintercept = mean.diff, color = "darkgoldenrod1", lwd = 1) + # mean
difference line
  geom_hline(yintercept = mean.diff + 1.96*sd.diff, color = "darkgoldenrod2",
lty = 2) + # upper 95% line
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
  geom_hline(yintercept = mean.diff - 1.96*sd.diff, color = "darkgoldenrod2",
lty = 2) + # lower 95% line
  # geom_smooth(method = "lm", level = 0.95, color = "black", lwd = 0.3) +
  labs(x = "Mean of filled/unfilled brain-predicted age (years)", y = "Unfilled
- Lesion-filled brain-predicted age (years)") +
  # ylim(c(-20,20)) +
  scale_fill_manual(values = ms.palette[-1]) +
  theme_bw() + theme(legend.position = c(0.1, 0.8)) +
  annotate("text", x = 25, y = mean.diff + 1.96*sd.diff + 0.5, label =
"+1.96*SD", color = "darkgoldenrod2") +
  annotate("text", x = 25, y = mean.diff - 1.96*sd.diff - 0.5, label = "-
1.96*SD", color = "darkgoldenrod2") +
  annotate("text", x = 27.5, y = mean.diff + 0.8, label = "Mean difference",
color = "darkgoldenrod2")
```

Warning: Removed 1441 rows containing missing values (geom_point).




```
ggsave("~/Work/Brain
ageing/Collaborations/MS/plots/lesion_filling_brain_age_BA_plot.pdf", width = 8,
height = 5, useDingbats = FALSE)
```

Warning: Removed 1441 rows containing missing values (geom_point).

LME model predicted brain-PAD based on subtype

This analysis excluded controls.

```
# LME model of subtype, including Cohort and scanner status as random effects
m2 <- lmer(BrainPAD ~ subtype + age_at_baseline_scan1 + gender + ICV +
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
(1|Cohort/scanner_status), data = subset(df.bl, df.bl$subtype != "control"))
summary(m2)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: BrainPAD ~ subtype + age_at_baseline_scan1 + gender + ICV + (1 |
##     Cohort/scanner_status)
##    Data: subset(df.bl, df.bl$subtype != "control")
##
## REML criterion at convergence: 8694.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.4362 -0.6514 -0.0882  0.5618  4.0050
##
## Random effects:
##  Groups              Name        Variance Std.Dev.
##  scanner_status:Cohort (Intercept)  9.784   3.128
##  Cohort               (Intercept)  1.227   1.108
##  Residual                         78.442   8.857
## Number of obs: 1204, groups:  scanner_status:Cohort, 19; Cohort, 15
##
## Fixed effects:
##                      Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)           8.00627    3.45908  936.99633   2.315   0.0209
## subtypeRRMS           6.08757    0.84400  841.57879   7.213 1.22e-12
## subtypeSPMS          11.63170    1.26777 1054.85471   9.175  < 2e-16
## subtypePPMS           6.10994    1.45708  455.22834   4.193 3.31e-05
## age_at_baseline_scan1  -0.16714    0.02866 1195.49837  -5.832 7.03e-09
## gendermale            1.42605    0.68186 1188.06560   2.091   0.0367
## ICV                   3.01383    2.33838 1186.29503   1.289   0.1977
##
## (Intercept)          *
## subtypeRRMS          ***
## subtypeSPMS          ***
## subtypePPMS          ***
## age_at_baseline_scan1 ***
## gendermale           *
## ICV
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) sbRRMS sbSPMS sbPPMS ag___1 gndrml
## subtypeRRMS -0.145
## subtypeSPMS -0.101  0.604
## subtypePPMS -0.093  0.516  0.489
## ag_t_bsln_1 -0.299 -0.112 -0.285 -0.246
## gendermale   0.498 -0.021 -0.080 -0.120  0.034
## ICV         -0.899 -0.005  0.057  0.032  0.001 -0.604

# round(confint(m2),2)
```

Brain-PAD estimated marginal means for subtypes

Generate EMMs for all MS subtypes. LME adjusting for age, gender, ICV, cohort and scanner status.

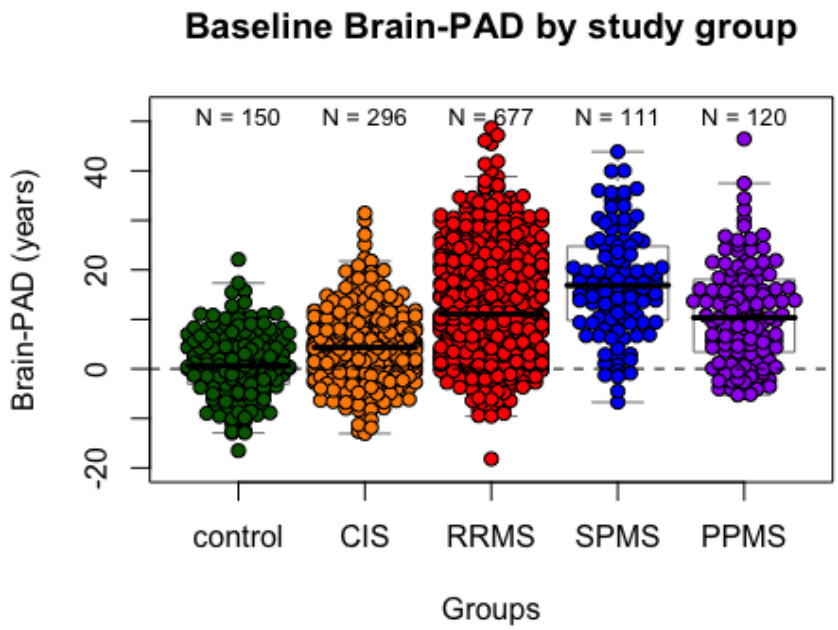
bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
fit <- (lmer(BrainPAD ~ subtype + age_at_baseline_scan1 + gender + ICV +
(1|Cohort/scanner_status), data = subset(df.bl, df.bl$subtype != "control")))
emmeans(object = fit, specs = c("subtype", "age_at_baseline_scan1", "ICV"))

##  subtype age_at_baseline_scan1     ICV    emmean       SE    df
##  CIS              39.4065 1.391797  6.327666 1.1878147 26.34
##  RRMS               39.4065 1.391797 12.415240 0.9835451 14.74
##  SPMS               39.4065 1.391797 17.959364 1.2869329 49.80
##  PPMS               39.4065 1.391797 12.437606 1.3751931 57.92
##   lower.CL  upper.CL
##   3.887597  8.767736
##  10.315633 14.514847
##  15.374229 20.544500
##   9.684780 15.190432
##
## Results are averaged over the levels of: gender
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
```

Brain-PAD boxplot by MS subtype

```
# calculate Ns
control_n <- with(df.bl, table(subtype))["control"]
cis_n <- with(df.bl, table(subtype))["CIS"]
rrms_n <- with(df.bl, table(subtype))["RRMS"]
spms_n <- with(df.bl, table(subtype))["SPMS"]
ppms_n <- with(df.bl, table(subtype))["PPMS"]

plot.pryr %<a-% {
with(df.bl, ehplot(BrainPAD, groups = subtype, ylim = c(-20,52), pch = 21, bg =
ms.palette[as.numeric(subtype)], box = T, offset = 0.1, intervals = 50, xlab =
"Groups", ylab = "Brain-PAD (years)", main = "Baseline Brain-PAD by study
group")) +
  abline(0,0, lty = 2) +
  text(1,51, paste("N = ", control_n, sep = ""), cex = 0.8) +
  text(2,51, paste("N = ", cis_n, sep = ""), cex = 0.8) +
  text(3,51, paste("N = ", rrms_n, sep = ""), cex = 0.8) +
  text(4,51, paste("N = ", spms_n, sep = ""), cex = 0.8) +
  text(5,51, paste("N = ", ppms_n, sep = ""), cex = 0.8)
}
plot.pryr
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



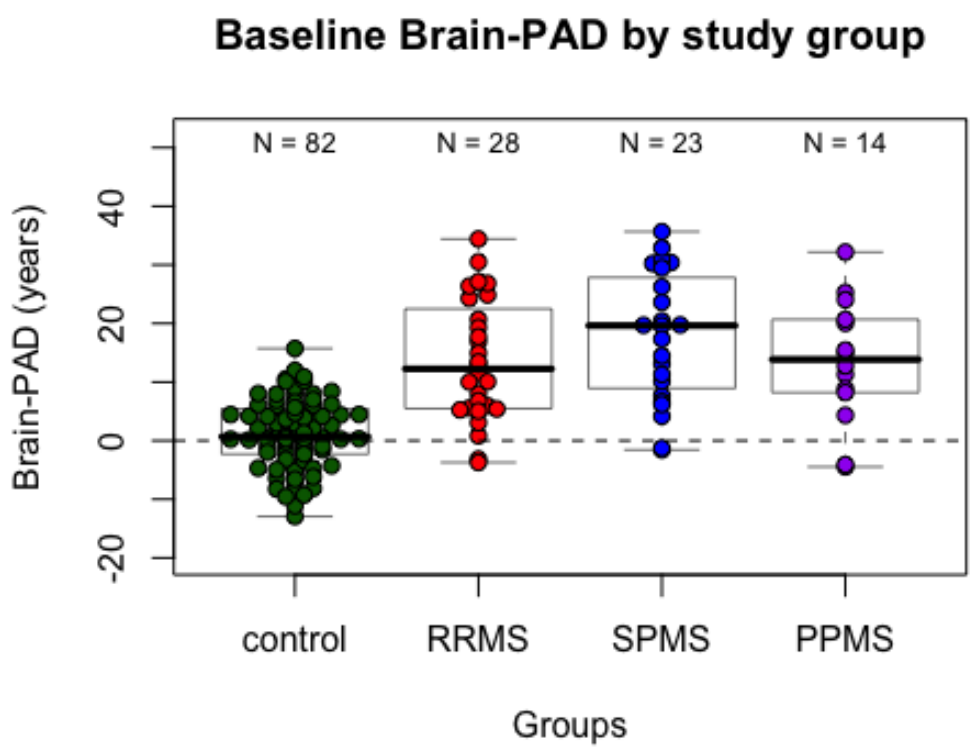

```
## integer(0)

cairo_pdf("plots/MS_UCL_MAGNIMS_combined_group_Brain-PAD.pdf", 8,8)
plot.pryr

## integer(0)

dev.off()

## quartz_off_screen
##           2
```

Brain-PAD boxplot by MS subtype in cohort UCL3 only

```
# calculate Ns
C3.bl <- df.bl %>% filter(Cohort == "UCL3")
C3.bl$subtype <- factor(C3.bl$subtype)
control_n <- with(C3.bl, table(subtype))["control"]
rrms_n <- with(C3.bl, table(subtype))["RRMS"]
spms_n <- with(C3.bl, table(subtype))["SPMS"]
ppms_n <- with(C3.bl, table(subtype))["PPMS"]

plot.pryr %<a-% {
  with(C3.bl, ehplot(BrainPAD, groups = subtype, ylim = c(-20,52), pch = 21, bg
= ms.palette[-2][as.numeric(subtype)], box = T, offset = 0.1, intervals = 50,
xlab = "Groups", ylab = "Brain-PAD (years)", main = "Baseline Brain-PAD by study
group")) +
  abline(0,0, lty = 2) +
  text(1,51, paste("N = ", control_n, sep = ""), cex = 0.8) +
  text(2,51, paste("N = ", rrms_n, sep = ""), cex = 0.8) +
  text(3,51, paste("N = ", spms_n, sep = ""), cex = 0.8) +
  text(4,51, paste("N = ", ppms_n, sep = ""), cex = 0.8)
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
}
plot.pryr
```




```
## integer(0)

cairo_pdf("plots/MS_cohort_C3_Brain-PAD.pdf", 8,8)
plot.pryr

## integer(0)

dev.off()

## quartz_off_screen
##                2
```

Post-hoc pairwise brain-PAD comparison of subtypes

```
round(with(df.bl, TukeyHSD(aov(BrainPAD ~ subtype)))$subtype, 4)

##                diff     lwr     upr  p adj
## CIS-control   3.4566  0.9312  5.9820 0.0018
## RRMS-control 11.3795  9.1056 13.6534 0.0000
## SPMS-control 16.1159 12.9611 19.2707 0.0000
## PPMS-control 10.0865  7.0005 13.1725 0.0000
## RRMS-CIS      7.9229  6.1671  9.6787 0.0000
## SPMS-CIS     12.6593  9.8548 15.4637 0.0000
## PPMS-CIS      6.6299  3.9030  9.3568 0.0000
## SPMS-RRMS     4.7364  2.1561  7.3166 0.0000
## PPMS-RRMS    -1.2930 -3.7888  1.2027 0.6180
## PPMS-SPMS    -6.0294 -9.3476 -2.7111 0.0000
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Brain-PAD by subtype descriptive statistics

```
with(df.bl, describeBy(BrainPAD, subtype, mat = T, digits = 1)) # brain-PAD by
MS patient subtypes and controls

##     item  group1 vars   n mean   sd median trimmed  mad   min  max range
## X11    1 control    1 150  1.1  6.7    0.6     1.2  6.5 -16.5 22.1  38.6
## X12    2     CIS    1 296  4.6  7.2    4.3     4.3  6.5 -13.1 31.5  44.5
## X13    3    RRMS    1 677 12.5 10.0   11.1    12.0  9.4 -18.2 48.7  66.8
## X14    4    SPMS    1 111 17.3 10.5   16.9    17.1 11.1  -6.7 43.9  50.6
## X15    5    PPMS    1 120 11.2 10.3   10.3    10.7 10.6  -5.3 46.4  51.7
##     skew kurtosis  se
## X11  0.0      0.1 0.6
## X12  0.5      0.9 0.4
## X13  0.5      0.3 0.4
## X14  0.2     -0.4 1.0
## X15  0.5      0.1 0.9
```

Correlates of brain-PAD at baseline

EDSS score, an index of disability

LME accounting for fixed effects of age at baseline, gender, ICV and random effects of Cohort and scanner status.

```
summary(lmer(BrainPAD ~ EDSSbaseline + age_at_baseline_scan1 + gender + ICV +
(1|Cohort/scanner_status), data = df.bl))

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: BrainPAD ~ EDSSbaseline + age_at_baseline_scan1 + gender + ICV +
##     (1 | Cohort/scanner_status)
##    Data: df.bl
##
## REML criterion at convergence: 8487.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7510 -0.6743 -0.0761  0.5761  3.6001
##
## Random effects:
##  Groups              Name        Variance Std.Dev.
##  scanner_status:Cohort (Intercept) 18.75    4.331
##  Cohort              (Intercept)  0.00    0.000
##  Residual                         78.02    8.833
## Number of obs: 1174, groups:  scanner_status:Cohort, 18; Cohort, 14
##
## Fixed effects:
##                      Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)           8.30507    3.52977  774.81179   2.353   0.0189
## EDSSbaseline          1.74180    0.18056 1133.59491   9.647  < 2e-16
## age_at_baseline_scan1  -0.20238    0.02916 1165.00045  -6.941 6.43e-12
## gendermale            1.03221    0.68309 1158.41082   1.511   0.1310
## ICV                   4.51312    2.35064 1156.85119   1.920   0.0551
##
## (Intercept)          *
## EDSSbaseline         ***
## age_at_baseline_scan1 ***
## gendermale
## ICV                  .
## ---
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) EDSSbs ag___1 gndrml
## EDSSbaselin -0.125
## ag_t_bsln_1 -0.279 -0.346
## gendermale   0.494 -0.107  0.031
## ICV         -0.894  0.107 -0.012 -0.605
## convergence code: 0
## singular fit
```

When predicting brain-PAD in a LME model, the effect of EDSS at baseline beta = 1.74, 95% CI = 1.39, 2.09, p = < 2.22e-16.

Test for interaction between subtype and EDSS on brain-PAD:

```
fit.edss <- lmer(BrainPAD ~ EDSSbaseline * subtype + age_at_baseline_scan1 +
gender + ICV + (1|Cohort/scanner_status), data = subset(df.bl, df.bl$subtype !=
"control"))
round(as.matrix(anova(fit.edss)["EDSSbaseline:subtype",]),3)

##                  Sum Sq Mean Sq NumDF    DenDF F value Pr(>F)
## EDSSbaseline:subtype 255.011  85.004     3 1159.588   1.125  0.338
```

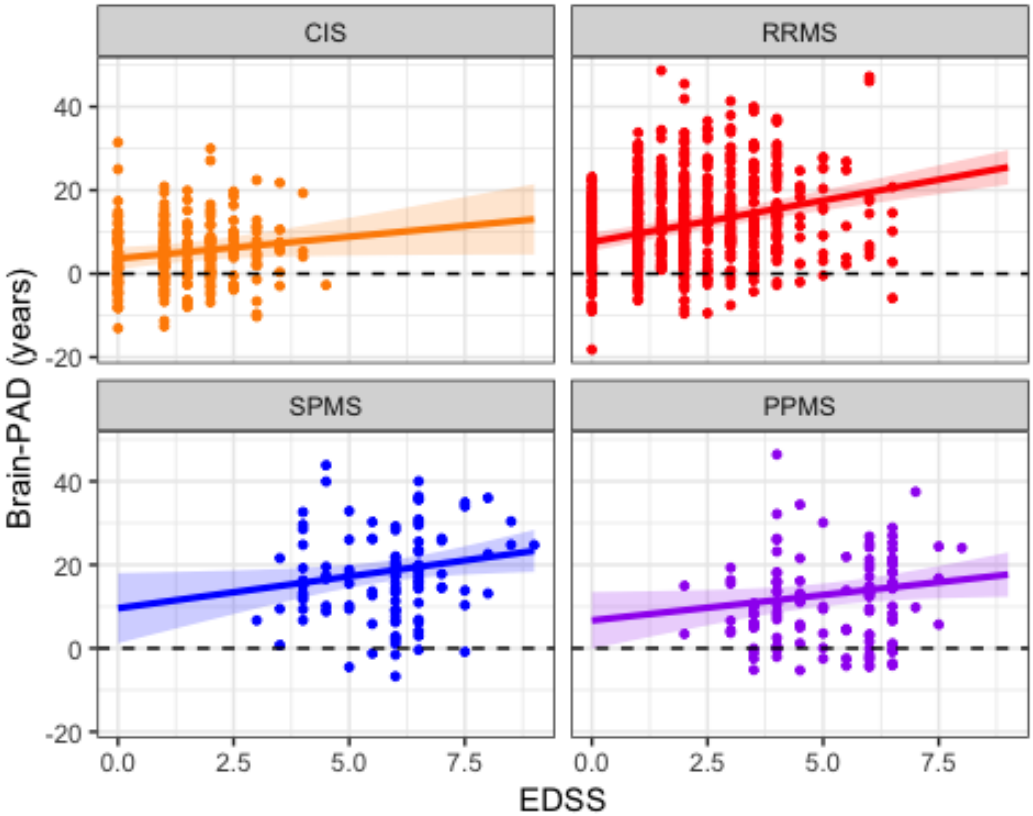
Use simple slopes from jtools to extract adjusted slopes for each subtype.

```
sim_slopes(fit.edss, pred = "EDSSbaseline", modx = "subtype", johnson_neyman =
F)

## SIMPLE SLOPES ANALYSIS
##
## Slope of EDSSbaseline when subtype = CIS:
##  Est. S.E.   df    p
##  1.04 0.54 1.93 0.05
##
## Slope of EDSSbaseline when subtype = RRMS:
##  Est. S.E.   df    p
##  1.99 0.28 7.22 0.00
##
## Slope of EDSSbaseline when subtype = SPMS:
##  Est. S.E.   df    p
##  1.53 0.70 2.18 0.03
##
## Slope of EDSSbaseline when subtype = PPMS:
##  Est. S.E.   df    p
##  1.22 0.61 2.00 0.05
```

Use interact_plot() from jtools to plot the adjusted slopes per group.

```
interact_plot(fit.edss, pred = "EDSSbaseline", modx = "subtype", plot.points =
T, interval = T, vary.lty = T, facet.modx = T, x.label = "EDSS", y.label =
"Brain-PAD (years)", point.size = 1, modx.labels = c("CIS", "RRMS", "SPMS",
"PPMS")) + geom_hline(yintercept = 0, lty = 2) + theme_bw()  +
  scale_fill_manual(values = ms.palette[2:5], name = "MS subtype") +
  scale_color_manual(values = ms.palette[2:5], name = "MS subtype")
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.




```
ggsave(filename = "~/Work/Brain ageing/Collaborations/MS/plots/EDSS_brain-PAD_plot.pdf", height = 5, width = 8, useDingbats = FALSE)
```

Age at diagnosis

LME accounting for fixed effects of age at baseline, gender, ICV and random effects of Cohort and scanner status. Exclude CIS patients and healthy controls.

```
summary(lmer(BrainPAD ~ disease_onset_age + age_at_baseline_scan1 + gender + ICV + (1|Cohort/scanner_status), data = subset(df.bl, df.bl$subtype != "control" & df.bl$subtype != "CIS")))

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: BrainPAD ~ disease_onset_age + age_at_baseline_scan1 + gender +
##     ICV + (1 | Cohort/scanner_status)
##    Data:
## subset(df.bl, df.bl$subtype != "control" & df.bl$subtype != "CIS")
##
## REML criterion at convergence: 6556.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2073 -0.6699 -0.0925  0.6307  3.8170
##
## Random effects:
##  Groups              Name        Variance Std.Dev.
##  scanner_status:Cohort (Intercept)  4.895   2.213
##  Cohort              (Intercept)  1.797   1.340
##  Residual                        83.165   9.119
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
## Number of obs: 900, groups:  scanner_status:Cohort, 18; Cohort, 14
##
## Fixed effects:
##                      Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)           16.73245    3.96818 850.54963   4.217 2.75e-05 ***
## disease_onset_age     -0.45261    0.04746 784.78256  -9.536  < 2e-16 ***
## age_at_baseline_scan1  0.16673    0.04480 775.76974   3.721 0.000213 ***
## gendermale             2.31995    0.78949 891.08283   2.939 0.003382 **
## ICV                    2.19226    2.72874 887.80921   0.803 0.421961
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) dss_n_ ag___1 gndrml
## dises_nst_g -0.020
## ag_t_bsln_1 -0.235 -0.720
## gendermale   0.491 -0.001 -0.014
## ICV         -0.922 -0.037  0.051 -0.585
```

When predicting brain-PAD in a LME model, the effect of age at diagnosis at baseline beta = -0.45, 95% CI = -0.55, -0.36, p = < 2.22e-16.

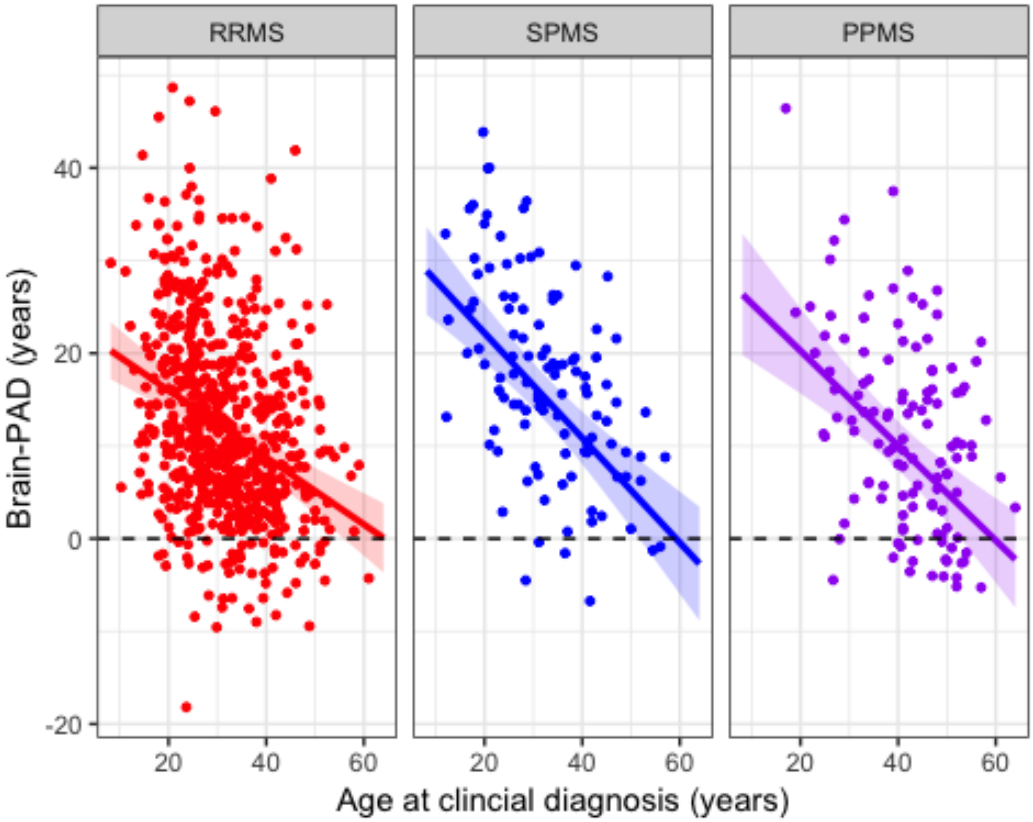
Test for interaction between subtype and age at diagnosis on brain-PAD:

```
fit.age <- lmer(BrainPAD ~ disease_onset_age * subtype + age_at_baseline_scan1 +
gender + ICV + (1|Cohort/scanner_status), data = subset(df.bl, df.bl$subtype !=
"control" & df.bl$subtype != "CIS"))
round(as.matrix(anova(fit.age)["disease_onset_age:subtype",]),3)

##                        Sum Sq Mean Sq NumDF   DenDF F value Pr(>F)
## disease_onset_age:subtype 524.445 262.222     2 883.928   3.203  0.041
```

Use simple slopes from jtools to extract adjusted slopes for each subtype.

```
sim_slopes(fit.age, pred = "disease_onset_age", modx = "subtype", johnson_neyman
= F)

## SIMPLE SLOPES ANALYSIS
##
## Slope of disease_onset_age when subtype = RRMS:
##   Est. S.E.    df    p
##  -0.36 0.05 -6.63 0.00
##
## Slope of disease_onset_age when subtype = SPMS:
##   Est. S.E.    df    p
##  -0.57 0.09 -6.40 0.00
##
## Slope of disease_onset_age when subtype = PPMS:
##   Est. S.E.    df    p
##  -0.51 0.10 -5.33 0.00

interact_plot(fit.age, pred = "disease_onset_age", modx = "subtype", plot.points
= T, interval = T, vary.lty = T, facet.modx = T, x.label = "Age at clincial
diagnosis (years)", y.label = "Brain-PAD (years)", point.size = 1, modx.labels =
c("RRMS", "SPMS", "PPMS")) + geom_hline(yintercept = 0, lty = 2) +  theme_bw()
+
  scale_fill_manual(values = ms.palette[3:5], name = "MS subtype") +
  scale_color_manual(values = ms.palette[3:5], name = "MS subtype")
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.




```
ggsave(filename = "~/Work/Brain
ageing/Collaborations/MS/plots/diagnosis_age_brain-PAD_plot.pdf", height = 5,
width = 8, useDingbats = FALSE)
```

Time since diagnosis

LME accounting for fixed effects of age at baseline, gender, ICV and random effects of Cohort and scanner status. Exclude controls, CIS patients and anyone with a time since diagnosis = 0.

```
summary(lmer(BrainPAD ~ disease_duration_at_baseline_scan1 +
age_at_baseline_scan1 + gender + ICV + (1|Cohort/scanner_status), data =
subset(df.bl, df.bl$subtype != "CIS" & df.bl$disease_duration > 0)))

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## BrainPAD ~ disease_duration_at_baseline_scan1 + age_at_baseline_scan1 +
##     gender + ICV + (1 | Cohort/scanner_status)
##    Data: subset(df.bl, df.bl$subtype != "CIS" & df.bl$disease_duration >
##     0)
##
## REML criterion at convergence: 6260.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1936 -0.6800 -0.0881  0.6373  3.5355
##
## Random effects:
##  Groups              Name        Variance Std.Dev.
##  scanner_status:Cohort (Intercept)  4.517   2.125
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
##  Cohort             (Intercept)  1.184   1.088
##  Residual                        85.098   9.225
## Number of obs: 857, groups:  scanner_status:Cohort, 18; Cohort, 14
##
## Fixed effects:
##                                 Estimate Std. Error        df t value
## (Intercept)                      17.15073    4.12903 828.28297   4.154
## disease_duration_at_baseline_scan1   0.44234    0.04813 743.77238   9.190
## age_at_baseline_scan1              -0.28569    0.03536 812.27062  -8.079
## gendermale                         2.19444    0.81213 848.75083   2.702
## ICV                              2.15264    2.84274 845.63208   0.757
##                                 Pr(>|t|)
## (Intercept)                      3.61e-05 ***
## disease_duration_at_baseline_scan1  < 2e-16 ***
## age_at_baseline_scan1            2.35e-15 ***
## gendermale                       0.00703 **
## ICV                              0.44912
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) d____1 ag___1 gndrml
## dss_drt___1  0.018
## ag_t_bsln_1 -0.334 -0.439
## gendermale   0.486  0.003 -0.019
## ICV         -0.927  0.036  0.023 -0.579
```

When predicting brain-PAD in a LME model, the effect of time since diagnosis at baseline beta = 0.48, 95% CI = 0.4, 0.57, p = < 2.22e-16.

Test for interaction between subtype and time since diagnosis on brain-PAD

```
fit.time <- lmer(BrainPAD ~ disease_duration_at_baseline_scan1 * subtype +
age_at_baseline_scan1 + gender + ICV + (1|Cohort/scanner_status), data =
subset(df.bl, df.bl$subtype != "control" & df.bl$subtype != "CIS" &
df.bl$disease_duration > 0))
round(as.matrix(anova(fit.time)["disease_duration_at_baseline_scan1:subtype",]),
3)
```

```
##                                  Sum Sq Mean Sq NumDF  DenDF
## disease_duration_at_baseline_scan1:subtype 603.971 301.986     2 690.45
##                                  F value Pr(>F)
## disease_duration_at_baseline_scan1:subtype   3.607  0.028
```

Use simple slopes from jtools to extract adjusted slopes for each subtype.

```
sim_slopes(fit.time, pred = "disease_duration_at_baseline_scan1", modx =
"subtype", johnson_neyman = F)
```

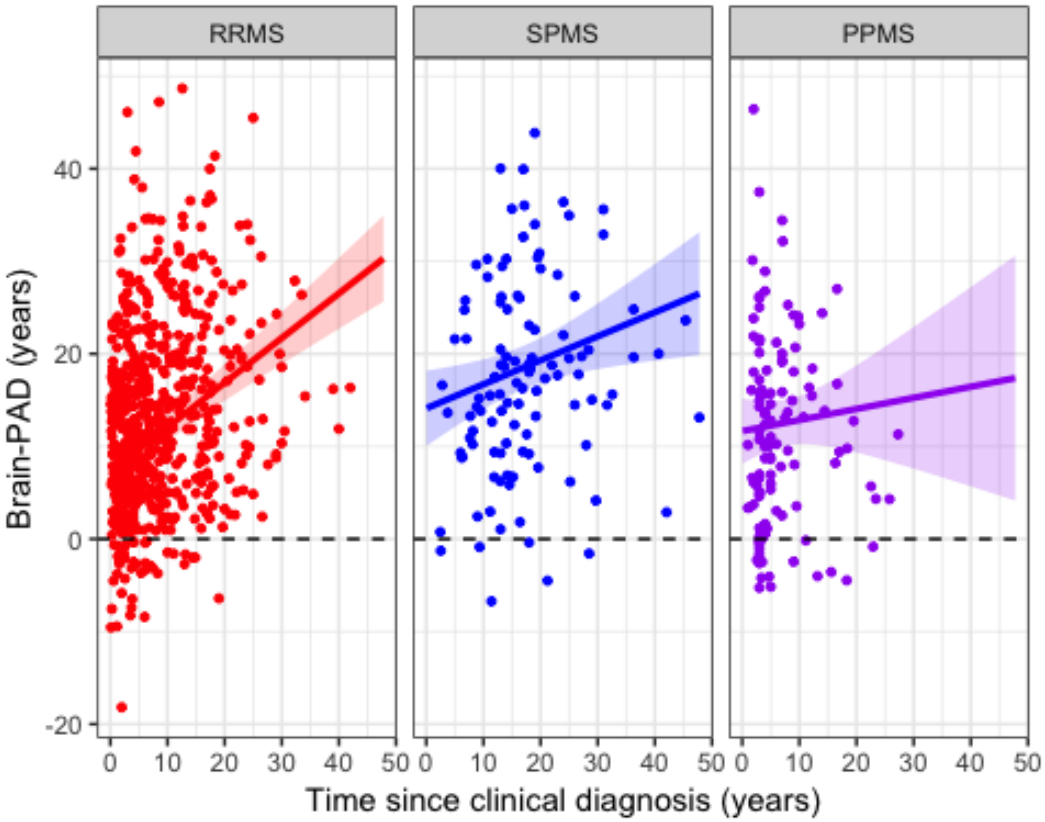
```
## SIMPLE SLOPES ANALYSIS
##
## Slope of disease_duration_at_baseline_scan1 when subtype = RRMS:
##  Est. S.E.   df    p
##  0.48 0.06 8.37 0.00
##
## Slope of disease_duration_at_baseline_scan1 when subtype = SPMS:
##  Est. S.E.   df    p
##  0.26 0.10 2.55 0.01
##
## Slope of disease_duration_at_baseline_scan1 when subtype = PPMS:
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
## Est. S.E.   df    p
## 0.12 0.17 0.73 0.47
```

Plot

```
interact_plot(fit.time, pred = "disease_duration_at_baseline_scan1", modx =
"subtype", plot.points = T, interval = T, vary.lty = T, facet.modx = T, x.label
= "Time since clinical diagnosis (years)", y.label = "Brain-PAD (years)",
point.size = 1, modx.labels = c("RRMS", "SPMS", "PPMS")) + geom_hline(yintercept
= 0, lty = 2) + theme_bw()  +
  scale_fill_manual(values = ms.palette[3:5], name = "MS subtype") +
  scale_color_manual(values = ms.palette[3:5], name = "MS subtype")
```




```
ggsave(filename = "~/Work/Brain
ageing/Collaborations/MS/plots/diagnosis_years_brain-PAD_plot.pdf", height = 5,
width = 8, useDingbats = FALSE)
```

EDSS progression survival analysis

Based on Arman Eshaghi's code used in Eshaghi et al., 2018 Annals of Neurology.

```
# function for characterising EDSS progression, based on different rates of
change and different baseline EDSS values
is_sustained_progression <- function(edssAtStart, change){
  sustainedProgression <- FALSE
  #if start of edss is 0, 1.5 increase is considered sustained progression
  if ((edssAtStart < 1) & (change >= 1.5)) {
    sustainedProgression <- TRUE
  }
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
  #if start of edss is 6 or above, 0.5 increase is considered sustained
progression
  else if ((edssAtStart >= 6) & (change >= 0.5 )) {
    sustainedProgression <- TRUE
  }
  #if start of edss is more than zero but less than 6, sustained progression is
by 1 increase in edss
  else if ((edssAtStart >= 1 ) & (edssAtStart < 6 )  & (change >= 1  )) {
    sustainedProgression <- TRUE
  }
  return(sustainedProgression)
}

## determine change in EDSS from baseline to last follow-up
## select latest EDSS per subject in subjects with 2 or more assessments
y1 <- df %>%
  filter(!subtype == "control")  %>%
  filter(!is.na(EDSSbaseline)) %>%
  group_by(PatientID) %>%
  top_n(1, interval) %>%
  ungroup() %>%
  dplyr::rename(latest_EDSS = EDSSatScan) %>%
  dplyr::select(PatientID, interval, EDSSbaseline, latest_EDSS) %>%
  filter(!is.na(latest_EDSS)) %>%
  mutate(EDSSchange = latest_EDSS - EDSSbaseline)

y1$EDSS_progression <-  mapply(is_sustained_progression, y1$EDSSbaseline,
y1$EDSSchange) # apply Arman's function
## get baseline brain-PAD and brain volumetric measures
y2 <- df %>%
  filter(!subtype == "control")  %>%
  filter(interval == 0) %>%
  filter(!is.na(EDSSbaseline)) %>%
  dplyr::rename(BrainPAD_baseline = BrainPAD) %>%
  dplyr::rename(GM_vol_baseline = GM_vol) %>%
  dplyr::rename(WM_vol_baseline = WM_vol) %>%
  dplyr::select(-one_of('interval'))

y3 <- right_join(y1, y2, by = c("PatientID")) %>%
  filter(!is.na(latest_EDSS))
```

Numbers of EDSS progressors

The number of MS patients with >= 2 EDSS scores was 1143.

```
table(y3$EDSS_progression) # calculate proportion of patients who progress

##
## FALSE  TRUE
##   840   303

round(prop.table(table(y3$EDSS_progression)),3) # calculate percentage of
patients who progress
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
##
## FALSE  TRUE
## 0.735 0.265
```

Run survival analysis

```
# creating new response function
S <- Surv(time = y3$interval, event = y3$EDSS_progression)
# Brain-PAD, age, sex model
m1 <- coxph(S ~ BrainPAD_baseline + age_at_baseline_scan1 + gender, data = y3)
summary(m1)

## Call:
## coxph(formula = S ~ BrainPAD_baseline + age_at_baseline_scan1 +
##     gender, data = y3)
##
##   n= 1143, number of events= 303
##
##                      coef exp(coef) se(coef)     z Pr(>|z|)
## BrainPAD_baseline     0.026343  1.026694 0.005534 4.761 1.93e-06 ***
## age_at_baseline_scan1 0.032452  1.032985 0.005750 5.644 1.66e-08 ***
## gendermale            0.217213  1.242609 0.119598 1.816   0.0693 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                     exp(coef) exp(-coef) lower .95 upper .95
## BrainPAD_baseline       1.027     0.9740     1.016     1.038
## age_at_baseline_scan1   1.033     0.9681     1.021     1.045
## gendermale              1.243     0.8048     0.983     1.571
##
## Concordance= 0.626  (se = 0.019 )
## Rsquare= 0.048   (max possible= 0.953 )
## Likelihood ratio test= 56.08  on 3 df,   p=4e-12
## Wald test            = 56.25  on 3 df,   p=4e-12
## Score (logrank) test = 57.15  on 3 df,   p=2e-12

cox.zph(m1)

##                      rho  chisq      p
## BrainPAD_baseline      0.0975 2.9164 0.0877
## age_at_baseline_scan1 -0.0826 2.0307 0.1541
## gendermale            -0.0122 0.0451 0.8319
## GLOBAL                     NA 5.3427 0.1484
```

The hazard ratio for brain-PAD on time-to-disease-progression was HR (95% CI) = 1.027, 1.016, 1.038. That means for every additional +1 year of brain-PAD there is a 1.027% increase in the likelihood of EDSS progression. Extrapolated over 5 years of brain-PAD, there is a 1.141 increase in the likelihood of EDSS progression.

Time-to-EDSS progression Kaplan-Meier plots

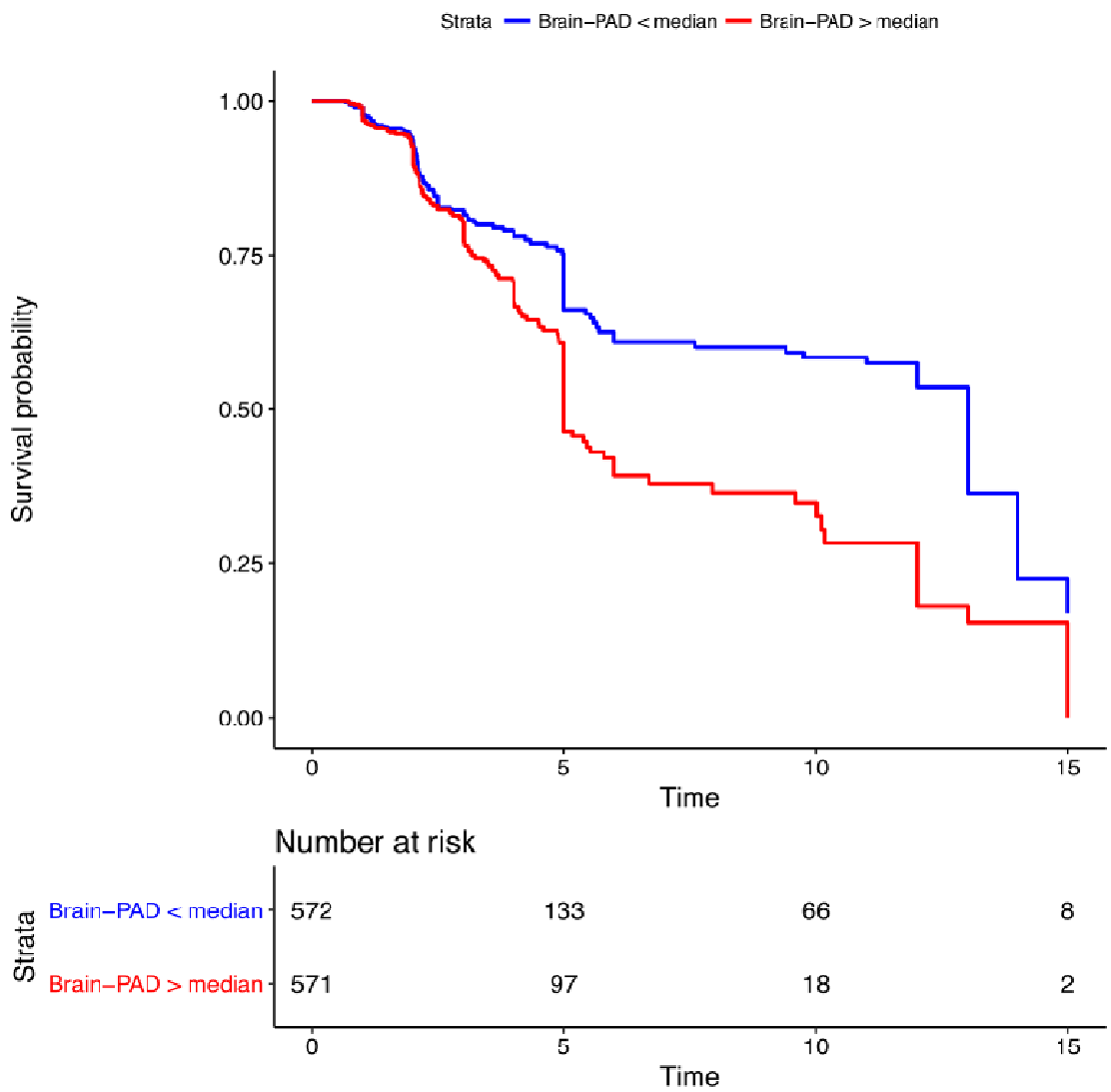
```
# Brain-PAD median split
km.plot.df <- y3 %>% mutate(split_BrainPAD = ntile(BrainPAD_baseline, 2))
```

Based on a median split of brain-PAD. The median value = 9.68 years.

```
# Run survplot on survival object
S <- Surv(time = km.plot.df$interval, event = km.plot.df$EDSS_progression) #
response function
survplot <- ggsurvplot(survfit(S ~ split_BrainPAD, data = km.plot.df),
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
surv.plot.height = 0.9, ggtheme = theme_survminer(), risk.table = T, cumcensor =
F, conf.int = F, palette = c("blue", "red"), censor = F , legend.labs =
c("Brain-PAD < median", "Brain-PAD > median"))
survplot
```




```
ggsave(filename = "~/Work/Brain ageing/Collaborations/MS/plots/KM_brain-
PAD_plot.pdf", height = 8, width = 8, print(survplot), useDingbats = FALSE)
```

Longitudinal brain-age analysis

The total number of people with two or more scans was n = 1266.

determine change in brain-PAD from baseline to last follow-up

```
## select latest brain-PAD per subject in subjects with 2 or more assessments
z1 <- df %>%
  filter(NoScans >= 2)  %>%
  group_by(PatientID) %>%
  top_n(1, interval) %>%
  ungroup() %>%
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
  dplyr::rename(latest_BrainPAD = BrainPAD) %>%
  dplyr::select(PatientID, interval, latest_BrainPAD)

## baseline brain-PAD
z2 <- df %>%
  filter(NoScans >= 2)  %>%
  filter(interval == 0) %>%
  filter(!is.na(BrainPAD)) %>%
  dplyr::rename(BrainPAD_baseline = BrainPAD) %>%
  dplyr::rename(GM_vol_baseline = GM_vol) %>%
  dplyr::rename(WM_vol_baseline = WM_vol) %>%
  dplyr::select(-one_of('interval'))

## calculate change in brain-PAD between baseline and latest brain-PAD
z3 <- right_join(z1, z2, by = c("PatientID")) %>%
  mutate(BrainPAD_change = latest_BrainPAD - BrainPAD_baseline)
```

Mean annualised rates of change in brain-PAD per group

```
describeBy(z3$BrainPAD_change/z3$interval, z3$subtype, mat = T, digits = 2)

##    item  group1 vars   n  mean   sd median trimmed  mad    min   max
## X11    1 control    1 111  0.03 2.02  -0.09   -0.12 1.14  -4.65  9.86
## X12    2     CIS    1 279  0.23 2.20   0.12    0.21 1.21 -11.74 11.49
## X13    3    RRMS    1 652  0.29 1.70   0.17    0.22 1.07 -12.24 16.21
## X14    4    SPMS    1 104 -0.29 1.60  -0.24   -0.27 0.93  -6.11  7.01
## X15    5    PPMS    1 119  0.40 1.52   0.30    0.40 1.38  -4.66  4.24
##    range  skew kurtosis   se
## X11 14.52  1.75     7.25 0.19
## X12 23.23  0.08     6.74 0.13
## X13 28.45  0.98    18.93 0.07
## X14 13.12  0.31     4.84 0.16
## X15  8.90 -0.18     0.62 0.14
```

determine change in EDSS from baseline to last follow-up

```
## select latest EDSS per subject in subjects with 2 or more assessments
a1 <- df %>%
  filter(NoScans >= 2)  %>%
  group_by(PatientID) %>%
  top_n(1, interval) %>%
  ungroup() %>%
  dplyr::rename(latest_EDSSatScan = EDSSatScan) %>%
  dplyr::select(PatientID, interval, latest_EDSSatScan)

## baseline EDSS
a2 <- df %>%
  filter(NoScans >= 2)  %>%
  filter(interval == 0) %>%
  filter(!is.na(EDSSatScan)) %>%
  dplyr::rename(EDSSatScan_baseline = EDSSatScan) %>%
  dplyr::rename(BrainPAD_baseline = BrainPAD) %>%
  dplyr::rename(GM_vol_baseline = GM_vol) %>%
  dplyr::rename(WM_vol_baseline = WM_vol) %>%
  dplyr::select(-one_of('interval'))

## calculate change in brain-PAD between baseline and latest brain-PAD
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
a3 <- right_join(a1, a2, by = c("PatientID")) %>%
  mutate(EDSS_change = latest_EDSSatScan - EDSSatScan_baseline)
```

Mean annualised rates of change in EDSS per group

```
with(subset(a3, a3$control0rest1 != "control"), describeBy(EDSS_change/interval,
subtype, mat = F, digits = 2))

##
##  Descriptive statistics by group
## group: control
## NULL
## ------------------------------------------------------------
## group: CIS
##    vars   n  mean   sd median trimmed mad   min max range  skew kurtosis
## X1    1 242 -0.26 1.05      0   -0.14 0.3 -6.86   3  9.86 -1.96     8.47
##      se
## X1 0.07
## ------------------------------------------------------------
## group: RRMS
##    vars   n mean   sd median trimmed  mad   min  max range skew kurtosis
## X1    1 635 0.12 0.45      0     0.1 0.19 -2.24 3.29  5.53 1.09    10.49
##      se
## X1 0.02
## ------------------------------------------------------------
## group: SPMS
##    vars   n mean   sd median trimmed  mad   min  max range skew kurtosis
## X1    1 104 0.14 0.29      0    0.11 0.07 -0.64 1.26   1.9 0.92     2.25
##      se
## X1 0.03
## ------------------------------------------------------------
## group: PPMS
##    vars   n mean   sd median trimmed  mad   min max range skew kurtosis
## X1    1 117 0.36 0.63   0.17    0.27 0.25 -1.01   3  4.01 1.92     5.35
##      se
## X1 0.06
```

Correlation between annualised EDSS change and brain-PAD change

```
delta.df <- join(a3, z3)
cor.test(delta.df$EDSS_change, delta.df$BrainPAD_change, method = "pearson")

##
##  Pearson's product-moment correlation
##
## data:  delta.df$EDSS_change and delta.df$BrainPAD_change
## t = 9.0089, df = 1095, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2067138 0.3169474
## sample estimates:
##       cor
## 0.2626875
```

Interaction between subtype and EDSS change

```
fit.change <- lm(BrainPAD_change ~ EDSS_change * subtype, data = delta.df)
anova(fit.change)

## Analysis of Variance Table
##
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
## Response: BrainPAD_change
##                    Df  Sum Sq Mean Sq F value    Pr(>F)
## EDSS_change         1  1452.1 1452.13 82.4760 < 2.2e-16 ***
## subtype             3   212.1   70.68  4.0146  0.007441 **
## EDSS_change:subtype 3   206.0   68.68  3.9008  0.008702 **
## Residuals        1089 19173.7   17.61
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

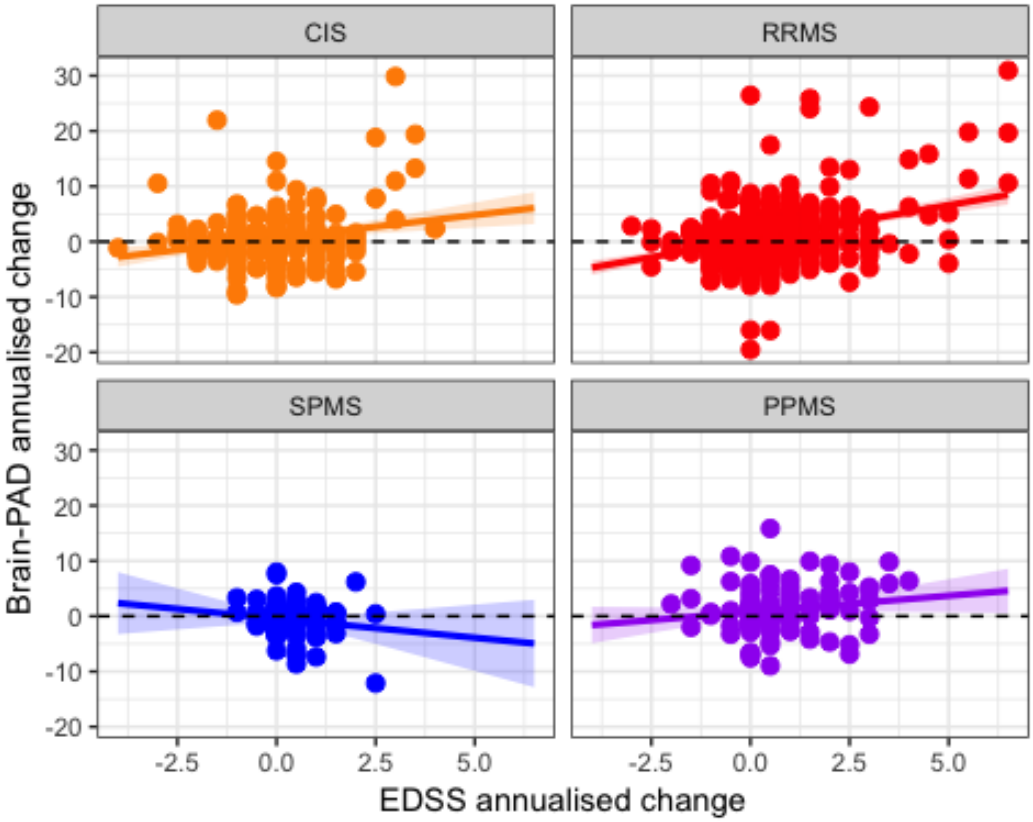
Use jtools package to get slopes from the model, per subtype.

```
sim_slopes(fit.change, pred = "EDSS_change", modx = "subtype", johnson_neyman =
F, digits = 4)

## SIMPLE SLOPES ANALYSIS
##
## Slope of EDSS_change when subtype = CIS:
##   Est.   S.E. t val.      p
##  0.8434 0.2189 3.8524 0.0001
##
## Slope of EDSS_change when subtype = RRMS:
##   Est.   S.E. t val.      p
##  1.2534 0.1459 8.5883 0.0000
##
## Slope of EDSS_change when subtype = SPMS:
##    Est.   S.E.  t val.      p
##  -0.6953 0.6510 -1.0680 0.2857
##
## Slope of EDSS_change when subtype = PPMS:
##   Est.   S.E. t val.      p
##  0.5882 0.3464 1.6983 0.0897
```

Plot

```
interact_plot(fit.change, pred = "EDSS_change", modx = "subtype", plot.points =
T, interval = T, facet.modx = T, x.label = "EDSS annualised change", y.label =
"Brain-PAD annualised change", modx.labels = c("CIS", "RRMS", "SPMS", "PPMS")) +
geom_hline(yintercept = 0, lty = 2) + theme_bw()  +
  scale_fill_manual(values = ms.palette[-1], name = "MS subtype") +
  scale_color_manual(values = ms.palette[-1], name = "MS subtype")
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.




```
ggsave(filename = "~/Work/Brain
ageing/Collaborations/MS/plots/change_EDSS_brain-PAD_plot.pdf", height = 5,
width = 8, useDingbats = FALSE)
```

Correlate baseline brain-PAD with the number of follow-up scans completed in the n=104 with >1 scan.

```
with(subset(df.bl, df.bl$NoScans > 1 & df.bl$subtype == "SPMS"),
cor.test(BrainPAD, NoScans, method = "spearman"))

## Warning in cor.test.default(BrainPAD, NoScans, method = "spearman"): Cannot
## compute exact p-value with ties

##
##  Spearman's rank correlation rho
##
## data:  BrainPAD and NoScans
## S = 241780, p-value = 0.002848
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.289771
```

Longitudinal brain-predicted age trajectories

Interaction between group and time

```
## conditional growth model - random effects of participant, cohort and scanner
status
model_int.group <- lmer(BrainPAD ~ control0rest1 * interval + gender +
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
age_at_baseline_scan1 + ICV + (interval| PatientID) + (1|Cohort/scanner_status),
data = df)
round(anova(model_int.group)["control0rest1:interval",],3)

## Type III Analysis of Variance Table with Satterthwaite's method
##                    Sum Sq Mean Sq NumDF  DenDF F value Pr(>F)
## control0rest1:interval 25.488  25.488     1 1325.6   5.374  0.021 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Brain-PAD change EMMs, using annualised difference between baseline and final follow-up.

Generate EMMs for all groups and MS subtypes. LME adjusting for age, gender, ICV, cohort and scanner status.

```
fit <- (lmer(BrainPAD_change ~ subtype + age_at_baseline_scan1 + gender + ICV +
(1|Cohort/scanner_status), data = z3))

## singular fit

emmeans(object = fit, specs = c("subtype"))

##  subtype     emmean       SE     df   lower.CL  upper.CL
##  control -0.9807015 0.5308630 102.29 -2.0336305 0.0722274
##  CIS     -0.1438444 0.4574657  34.93 -1.0726189 0.7849301
##  RRMS     0.9347867 0.3421557  16.77  0.2121440 1.6574293
##  SPMS     0.3405131 0.5200587 102.29 -0.6909864 1.3720125
##  PPMS     1.2072129 0.5259459  90.74  0.1624462 2.2519795
##
## Results are averaged over the levels of: gender
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
```

Generate EMMs for HCs and MS/CIS combined. LME adjusting for age, gender, ICV, cohort and scanner status.

```
fit <- (lmer(BrainPAD_change ~ control0rest1 + age_at_baseline_scan1 + gender +
ICV + (1|Cohort/scanner_status), data = z3))
emmeans(object = fit, specs = c("control0rest1"))

##  control0rest1     emmean       SE     df   lower.CL   upper.CL
##  control      -1.0983087 0.5298190 108.79 -2.1484159 -0.0482016
##  MS            0.6999868 0.3188504  12.36  0.0075069  1.3924667
##
## Results are averaged over the levels of: gender
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
```

Slopes per group

Controls

```
control.slope <- summary(lmer(BrainPAD ~ interval + gender +
age_at_baseline_scan1 + ICV + (interval| PatientID) + (1|Cohort/scanner_status),
data = subset(df, df$control0rest1 == "control")))$coef["interval","Estimate"]
control.intercept <- summary(lmer(BrainPAD ~ interval + gender +
age_at_baseline_scan1 + ICV + (interval| PatientID) + (1|Cohort/scanner_status),
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
data = subset(df, df$control0rest1 ==
"control")))$coef["(Intercept)","Estimate"]
```

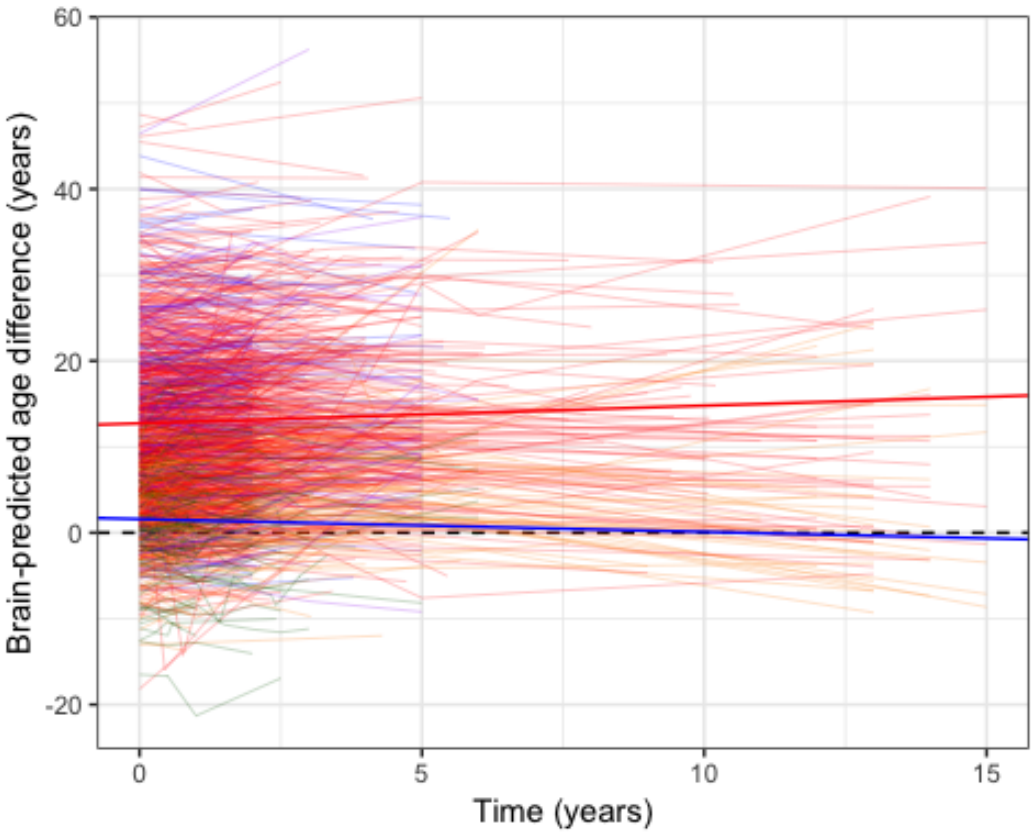
MS patients

```
ms.slope <- summary(lmer(BrainPAD ~ interval + gender + age_at_baseline_scan1 +
ICV + (interval| PatientID) + (1|Cohort/scanner_status), data = subset(df,
df$control0rest1 == "MS")))$coef["interval","Estimate"]
ms.intercept <- summary(lmer(BrainPAD ~ interval + gender +
age_at_baseline_scan1 + ICV + (interval| PatientID) + (1|Cohort/scanner_status),
data = subset(df, df$control0rest1 == "MS")))$coef["(Intercept)","Estimate"]
```

Longitudinal brain-PAD by interval plots

```
## subtype longindtual slopes
# intercept <- summary(model_int)$coefficients["(Intercept)","Estimate"]
# cis.slope <- summary(model_int)$coefficients["subtypeCIS:interval","Estimate"]
# rrms.slope <-
summary(model_int)$coefficients["subtypeRRMS:interval","Estimate"]
# spms.slope <-
summary(model_int)$coefficients["subtypeSPMS:interval","Estimate"]
# ppms.slope <-
summary(model_int)$coefficients["subtypePPMS:interval","Estimate"]
# plot.df <- data.frame(intercept = intercept, slope = c(cis.slope, rrms.slope,
spms.slope, ppms.slope), subtype = c("CIS", "RRMS", "SPMS", "PPMS"))

ggplot(data = df, aes(x = interval, y = BrainPAD, fill = subtype)) +
  geom_hline(yintercept = 0, lty = 2) +
  geom_line(aes(group = PatientID, colour = subtype), alpha = 0.3, linetype = 1,
size = 0.25) +
  # geom_point(aes(fill = subtype), pch = 21, size = 0.5, alpha = 0.2) +
  geom_abline(slope = control.slope, intercept = control.intercept, col =
"blue") +
  geom_abline(slope = ms.slope, intercept = ms.intercept, col = "red") +
  labs(x = "Time (years)", y = "Brain-predicted age difference (years)") +
  scale_fill_manual(values = ms.palette) +
  scale_color_manual(values = ms.palette) +
  # facet_wrap(~ subtype, scales = "free_x") +
  theme_bw() + theme(legend.position = "none")

## Warning: Removed 1 rows containing missing values (geom_path).
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.




```
ggsave(filename = "~/Work/Brain
ageing/Collaborations/MS/plots/longitudinal_brain-PAD_time_plot.pdf", height =
6, width = 6, useDingbats = FALSE)

## Warning: Removed 1 rows containing missing values (geom_path).
```

Supplementary analysis

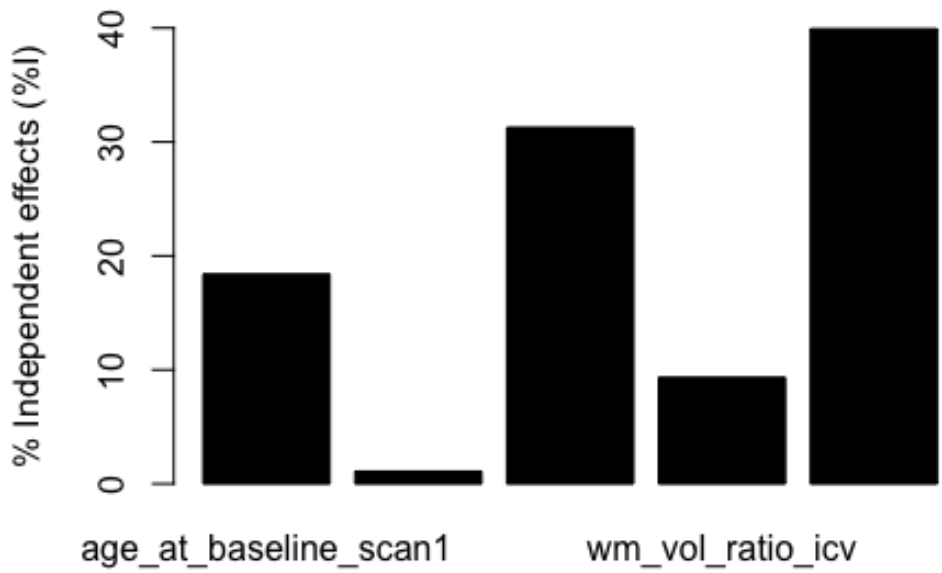
Hierarchical partitioning of brain-PAD

```
a <- summary(lm(BrainPAD ~ age_at_baseline_scan1 + gender + gm_vol_ratio_icv +
wm_vol_ratio_icv + csf_vol_ratio_icv, data = df.bl))
print(a)

##
## Call:
## lm(formula = BrainPAD ~ age_at_baseline_scan1 + gender + gm_vol_ratio_icv +
##     wm_vol_ratio_icv + csf_vol_ratio_icv, data = df.bl)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -39.527  -4.842   0.190   4.790  29.445
##
## Coefficients: (1 not defined because of singularities)
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           156.78460    4.26998   36.72  < 2e-16 ***
## age_at_baseline_scan1   -0.46373    0.02408  -19.26  < 2e-16 ***
## gendermale             -2.01799    0.43216   -4.67 3.32e-06 ***
## gm_vol_ratio_icv      -196.06829    6.02802  -32.53  < 2e-16 ***
## wm_vol_ratio_icv      -111.19551    6.98622  -15.92  < 2e-16 ***
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
## csf_vol_ratio_icv          NA       NA     NA     NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.433 on 1349 degrees of freedom
## Multiple R-squared:  0.4877, Adjusted R-squared:  0.4862
## F-statistic:   321 on 4 and 1349 DF,  p-value: < 2.2e-16
```

```
h.p <- hier.part(y = df.bl$BrainPAD, xcan =
df.bl[c("age_at_baseline_scan1","gender","gm_vol_ratio_icv", "wm_vol_ratio_icv",
"csf_vol_ratio_icv")], gof = "Rsqu")
```




```
print(h.p)

## $gfs
##  [1] 0.0000000000 0.0001847944 0.0054708411 0.2503255424 0.0805476003
##  [6] 0.3448462682 0.0056460035 0.3845513168 0.0810550023 0.4499544177
## [11] 0.2510562177 0.0854193771 0.3458685152 0.3456355219 0.3456355219
## [16] 0.3456355219 0.3914809503 0.0859101977 0.4541394054 0.4794083693
## [21] 0.4794083693 0.4794083693 0.3468458772 0.3468458772 0.3468458772
## [26] 0.3456355219 0.4876890065 0.4876890065 0.4876890065 0.4794083693
## [31] 0.3468458772 0.4876890065
##
## $IJ
##                          I           J       Total
## age_at_baseline_scan1 0.089705020 -0.0895202260 0.0001847944
## gender                0.005310567  0.0001602739 0.0054708411
## gm_vol_ratio_icv      0.152491187  0.0978343555 0.2503255424
## wm_vol_ratio_icv      0.045488042  0.0350595578 0.0805476003
## csf_vol_ratio_icv     0.194694190  0.1501520787 0.3448462682
```

bioRxiv preprint doi: https://doi.org/10.1101/584888; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


```
##
## $I.perc
##                        I
## age_at_baseline_scan1 18.393898
## gender                1.088925
## gm_vol_ratio_icv      31.268121
## wm_vol_ratio_icv       9.327264
## csf_vol_ratio_icv     39.921792

sum(h.p$IJ["I"]) - a$adj.r.squared

## [1] 0.001519084

sum(h.p$IJ["I"])

## [1] 0.487689

sum(h.p$IJ["J"])

## [1] 0.193686

round(h.p$IJ, 3)

##                      I     J Total
## age_at_baseline_scan1 0.090 -0.090 0.000
## gender                0.005  0.000 0.005
## gm_vol_ratio_icv      0.152  0.098 0.250
## wm_vol_ratio_icv      0.045  0.035 0.081
## csf_vol_ratio_icv     0.195  0.150 0.345

round(h.p$I.perc, 1)

##                     I
## age_at_baseline_scan1 18.4
## gender                1.1
## gm_vol_ratio_icv      31.3
## wm_vol_ratio_icv       9.3
## csf_vol_ratio_icv     39.9
```