JCBFM

Progression of MRI markers in cerebral small vessel disease: Sample size considerations for clinical trials

Philip Benjamin[1],*, Eva Zeestraten[1],*, Christian Lambert[1], Irina Chis Ster[2], Owen A Williams[1], Andrew J Lawrence[3], Bhavini Patel[1], Andrew D MacKinnon[4], Thomas R Barrick[1] and Hugh S Markus[3]

Abstract

Detecting treatment efficacy using cognitive change in trials of cerebral small vessel disease (SVD) has been challenging, making the use of surrogate markers such as magnetic resonance imaging (MRI) attractive. We determined the sensitivity of MRI to change in SVD and used this information to calculate sample size estimates for a clinical trial. Data from the prospective SCANS (St George's Cognition and Neuroimaging in Stroke) study of patients with symptomatic lacunar stroke and confluent leukoaraiosis was used ($n = 121$). Ninety-nine subjects returned at one or more time points. Multimodal MRI and neuropsychologic testing was performed annually over 3 years. We evaluated the change in brain volume, T2 white matter hyperintensity (WMH) volume, lacunes, and white matter damage on diffusion tensor imaging (DTI). Over 3 years, change was detectable in all MRI markers but not in cognitive measures. WMH volume and DTI parameters were most sensitive to change and therefore had the smallest sample size estimates. MRI markers, particularly WMH volume and DTI parameters, are more sensitive to SVD progression over short time periods than cognition. These markers could significantly reduce the size of trials to screen treatments for efficacy in SVD, although further validation from longitudinal and intervention studies is required.

Introduction

Cerebral small vessel disease (SVD) is a term used to describe a group of pathologic processes that affect the perforating cerebral arterioles and capillaries resulting in brain injury to the subcortical grey and white matter.[1] A number of brain parenchymal pathologies can occur, including small deep infarcts, microbleeds, areas of diffuse gliosis, ischemic demyelination and axonal loss corresponding to regions of radiologic leukoaraiosis, and diffuse brain atrophy.[2] Clinically SVD presents with lacunar strokes, which represent approximately 20% of all ischemic strokes, and it is also the major cause of vascular cognitive impairment. In addition, it appears to interact with Alzheimer's disease, exacerbating the degree of clinical impairment.[3]

[1]Neurosciences Research Centre, Cardiovascular and Cell Sciences Research Institute, St George's University of London, London, UK
[2]Institute of Infection and Immunity, St George's University of London, London, UK
[3]Clinical Neurosciences, University of Cambridge, Cambridge, UK
[4]Atkinson Morley Regional Neuroscience Centre, St George's NHS Healthcare Trust, London, UK

*These authors contributed equally to this work.

Corresponding author:
Philip Benjamin, Stroke and Dementia Research Centre, St George's University of London, Cranmer Terrace, Tooting, London SW17 0RE, UK.
Email: philipbenjamin@doctors.net.uk

Thus SVD is an enormous health burden that leads to significant neurologic and cognitive decline. Despite its importance, there are few validated treatments for SVD. Reasons for this include a lack of understating of the pathogenesis and the challenges of performing large clinical trials in the condition including the insensitivity of cognitive scores to change.[4] The use of surrogate markers in which novel treatments can be evaluated for potential efficacy before large clinical end point trails is attractive and magnetic resonance imaging (MRI) has been suggested as such a surrogate.

MRI is essential to diagnosis of SVD. Common features seen on conventional MRI include lacunes, T2 white matter hyperintensities (WMHs), cerebral microbleeds (CMBs), perivascular spaces, and brain atrophy.[2] More advanced techniques using diffusion tensor imaging (DTI) has been shown to be sensitive to tissue damage showing abnormalities in apparently normal appearing white matter.[5,6] These abnormalities on MRI may be potentially useful surrogate disease markers that can used to assess therapeutic approaches. If these MRI parameters are to be used as reliable surrogate markers in clinical trials, they must fulfill the following criteria: (1) they must be able to predict clinical outcome, as changes induced by a therapy on a surrogate marker are expected to reflect changes in a clinically meaningful end point; (2) change in a surrogate marker must be detectable prospectively; and (3) the sample size required to show therapeutic efficacy should be feasible in the setting of a clinical trial.[7] The sample size is dependent on a number of factors, including the magnitude of the change in the marker detectable prospectively and the variability of measurements.

Several potential MRI markers in SVD are known to predict changes in a clinical outcome measure (e.g., cognitive impairment, progression to dementia, or disability; Table 1). Whole brain volume has been shown to correlate strongly with cognitive impairment in several cross-sectional[8,9] and longitudinal[10] studies. Brain volume change has also been shown to correlate with disability scales in Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leucoencephalopathy (CADASIL), a genetic form of SVD.[11] T2 WMHs, have usually shown significant associations with cognitive impairment cross-sectionally[8] and longitudinally,[4] although the correlation has been weaker or sometimes absent in subjects with symptomatic SVD and leukoaraiosis[8,9] (Table 1). DTI parameters in normal-appearing white matter have consistently been shown to correlate with cognitive performance in a number of SVD cohorts.[5,12,13] Finally, lacunes are also an important predictor of cognitive impairment in SVD.[14] Patients with incident lacunes are known to have a faster rate of cognitive decline.[15]

Abnormalities on brain imaging often exist long before symptoms occur and may have a faster rate of progression than cognitive outcome measures. Using them as surrogate markers could significantly reduce the size, duration, and cost of clinical trials in SVD. In this study, we use multimodal serial brain (MRI) to prospectively evaluate the rate of change of some of these proposed MRI markers in established SVD; in particular brain atrophy, T2 WMH, lacunes and DTI parameters. We then used this data to determine the sample sizes required to demonstrate a reduction in the rate of progression of these MR abnormalities in a hypothetical clinical trial and compared these estimates with estimates if cognitive measures were used.

Materials and Methods

Patients

Patients with SVD were recruited as part of the prospective St George's Cognition and Neuroimaging in Stroke (SCANS) study.[8] Recruitment was from acute stroke units or outpatient stroke clinics in three hospitals covering a contiguous catchment area in South London (St George's, King's College, and St Thomas' Hospitals). Inclusion criteria comprised a clinical lacunar stroke syndrome[16] with an anatomically corresponding lacunar infarct on MRI in addition to confluent WMH on MRI (Fazekas grade 2).[17] Exclusion criteria were: any cause of stroke mechanism other than SVD (e.g., cardioembolic source or extra- or intra-cerebral artery stenosis of >50%), other major central nervous system disorders, major psychiatric disorders, any other cause of white matter disease, contra-indications to MRI, or non-fluent in English. The study was approved by Wandsworth research ethics committee, and all patients gave written informed consent. MRI acquisitions and cognitive assessments were performed at least 3 months after the last stroke to exclude acute effects on cognition. All patients were also screened for cardiovascular risk factors, including hypertension (defined as systolic blood pressure >140 mmHg or diastolic >90 mmHg or treatment with antihypertensive drugs), hypercholesterolemia (defined as a serum total cholesterol >5.2 mmol/l or treatment with a statin), diabetes mellitus, and smoking.

Subjects were invited back annually for cognitive testing and MRI scanning. In this analysis, follow-up data up to the third follow-up was used, providing a maximum of four data points per individual.

MRI Acquisition

Images were acquired on a 1.5 T Signa HDxt scanner (General Electric, Milwaukee, WI, USA) with

Table 1. Longitudinal studies investigating MRI markers of SVD and their relationship to clinical measures.

Study	Participants	N (follow-up period)	Clinical tests	MRI measure	Findings
T2 white matter hyperintensities					
Garde et al.[41]	Danish heathy subjects >50 years	698 (3–5 years)	Wechsler adult intelligence scale	Automated quantitative	Increase in WMH volume was correlated with a decline in verbal IQ
Holtmann-spötter et al.[5]	CADASIL	62 (23–31 moths)	Rankin (disability) scale, Barthel index, global cognitive function	Semi-automated quantitative	No association between change in WMH and clinical scores
Van den Heuvel et al.[42]	PROSPER study (elderly)	554 (3 years)	Mental processing speed	Semi-automated quantitative	Progression of periventricular WMH was associated with a decline in processing speed
Liem et al.[33]	CADASIL	25+13 controls (7 years)	Global cognitive function, memory, executive function, processing speed, language, intelligence	Semi-automated quantitative	WMH were not associated with cognitive decline
Schmidt et al.[4]	LADIS study	394 (3 years)	Vascular dementia assessment scale	Visual rating scale (Rotterdam progression scale)	WMH progression was related to deterioration in cognitive function
Brain volume					
Schmidt et al.[43]	Austrian Stroke Prevention Study	329 (6 years)	Memory, learning abilities, conceptional reasoning, attention, speed, visuopractical skills	Automated quantitative (SIENA)	Brain volume loss was the strongest predictor of decline in mnestic, visuopractical and attention/speed performance
Peters et al.[11]	CADASIL	76 (2 years)	Rankin (disability) scale, Barthel index, global cognitive function	Automated quantitative	Brain volume change significantly correlated with all measures of disability and global cognitive functioning

Table 1. Continued

Study	Participants	N (follow-up period)	Clinical tests	MRI measure	Findings
Jokinen et al.[10]	LADIS study	477 (3 years)	MMSE, VADAS, processing speed, executive functions, memory	Visual template-based visual rating scale	Global atrophy predicted decline in MMSE, VADAS, speed and executive functions
Diffusion tensor imaging					
Holtmannspötter et al.[5]	CADASIL	62 (23–31 months)	Rankin (disability) scale, Barthel index, global cognitive function	Mean MD	The mean MD change correlated significantly with deterioration of physical disability and global cognitive function
Charlton et al.[34]	Elderly subjects	84 (2 years)	Executive function, working memory, information processing speed	Median MD	Change in MD median was associated with worsening working memory
Lacunes					
Jokinen et al.[15]	LADIS study	387 (3 years)	Executive function, processing speed, global cognitive function	Number of new lacunes	Incident lacunes on MRI parallel a steeper rate of decline in executive functions and psychomotor speed

Abbreviations: CADASIL, Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leucoencephalopathy; IQ, intelligent quotient; LADIS, Leukokraurosis and Disability; MD, mean diffusivity; MMSE, Mini Mental State Examination; MRI, magnetic resonance imaging; SIENA, Structural Image Evaluation using Normalisation of Atrophy; SVD, small vessel disease; VADAS, Vascular Dementia Assessment Scale; WMH, white matter hyperintensity.

maximum gradient amplitude of 33 mT/m and a proprietary head coil. All image sequences were acquired across the whole brain and total imaging time was approximately 45 minutes. Patients were placed in the head coil in a neutral position with an alignment marker at the nasal bridge to standardize head position. Minimal movement was ensured by use of foam pads and a velcro strap across the forehead.

The imaging protocol included: (1) Fluid Attenuated Inversion Recovery (FLAIR) sequence—TR/TE/TI = 9,000/130/2,200 ms, field-of-view (FOV) = 240 × 240 mm^2, matrix = 256 × 192, 28 axial slices of 5-mm thickness. (2) Spoiled gradient echo recalled T1-weighted (SPGR) 3D coronal sequence—TR/TE = 11.5/5 ms, FOV = 240 × 240 mm^2, matrix = 256 × 192, flip angle = 18°, 176 coronal slices of 1.1-mm thickness reconstructed to an in plane resolution of 1.1 mm. (3) DTI—Axial single shot spin echo planar images (EPI; TE = 93.4 ms, TR = 15,600 ms) were acquired to achieve whole brain coverage (2.5 mm^3 isotropic voxels; FOV = 240 × 240 mm^2, acquisition matrix = 96 × 96), following four acquisitions without diffusion weighting ($b = 0$ s/mm), diffusion-sensitized images were acquired with gradients applied ($b = 1000$ s/mm) in 25 non-collinear directions. This was repeated to acquire a further four $b = 0$ s/mm images and the negative of the 25 directions. Diffusion-weighted images were realigned to remove eddy current distortions using the FSL Linear Image Registration Tool (FLIRT, FMRI Software Library, FSL version 4.1; FMRIB Analysis Group, Oxford, UK, http://www.fmrib.ox.ac.uk/fsl). The geometric average of the positive and negative acquisitions was taken to eliminate gradient cross-terms. The eight $b = 0$ s/mm images were co-registered and averaged to give a T2-weighted echo planar image (henceforth termed b0).

Estimation of Longitudinal Brain Volume Change

Brain volume at baseline was calculated using a fully automated program, SIENAX (Cross-sectional Structural Image Evaluation using Normalisation of Atrophy; www.fmrib.ox.ac.uk/fsl) on T1-weighted images. This program computes brain volume relative to normalized skull size. Longitudinal changes in global brain volume were estimated from the T1-weighted images for each subject using an automated image analysis tool SIENA (Structural Image Evaluation using Normalisation of Atrophy; www.fmrib.ox.ac.uk/fsl).[18] This tool was used as SVD is a diffuse neurodegenerative disease and SIENA provides a sensitive measure of global volume loss. SIENA brain volume change has also been shown to be a feasible outcome measure in clinical trials in multiple sclerosis[19] and Alzheimer's disease.[20] SIENA estimates the percentage of brain volume change (PBVC) between two input images and was applied here to provide a global estimate of PBVC between each time point. Brain volume at each time point was then calculated using the baseline brain volume and the PBVC between each time point.

Diffusion MRI Analysis

Diffusion tensor elements were fitted at each voxel using DTIfit in FSL www.fmrib.ox.ac.uk/fsl, and mean diffusivity (MD) and fractional anisotropy (FA) maps were calculated.

Diffusion parameter histograms were then obtained in white matter using the following method. T1-weighted images were segmented using New Segment in SPM12 into isotropic (1 mm^3 voxel resolution) grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF) tissue classes. These segmentations were used to estimate deformations to a group-average template generated from the data cohort using the Shoot toolbox in SPM12.[21] The T1-weighted and FLAIR images were skull-stripped using the segmentations at a threshold of 0.1 and warped to the common group-average template. These were used to create population-specific Tissue Probability Maps (TPMs) using a modified multivariate mixture of Gaussians.[22] This technique increases segmentation accuracy across the cohort and provides TPMs for WMH from the FLAIR and T1-weighted images, allowing automatic segmentation at an individual level. These TPMs were used to replace the default SPM TPMs, and New Segment was re-run on the native space images to generate improved segmentations for GM, WM, and CSF, while also providing a WMH tissue class. The WMH segmentation maps were binarized at an individual set threshold by checking each of the results manually to ensure accurate correspondence. The binarization threshold was performed without referring to previous time points. Results were manually refined where necessary to optimize accuracy, with good intra- and inter-rater reliability (Pearson's intraclass correlation coefficient 0.98 and 0.99, respectively, defined on a randomly selected group of 20 scans between two raters (CL and EZ)). An additional step was performed to repair the segmentations for regions of tissue damage. This was motivated by the observation that regions affected by pathology are frequently misclassified during routine segmentation. For example, regions of gliosis are misclassified as CSF and regions of WMH as GM leading to erroneous deformation estimations and subsequent inaccuracies in estimating warps to the group-average space. For this reason, WMH regions were automatically reclassified as WM and any erroneously classified tissue was removed from the GM and CSF segmentation maps.[14]

Affine transformations, followed by non-linear transformations (using the FMRIBs Non-linear Image Registration Tool (FNIRT)) were computed by registration of the T1-weighted images to the $b0$ image of each subject and applied to the above segmentation maps to obtain masks of brain tissue in $b0$ space. Voxels were included in the masks if the probability of being GM or WM or WMH was >0.5. CSF voxels were removed using thresholds based on the diffusion properties of CSF (MD values of >0.0026 mm^2/s were considered to contain CSF and excluded from the analysis). The diffusion characteristics of the WM tissue class were evaluated for the purposes of this study.

MD and FA histograms in voxels representing WM were calculated. For each subject, a histogram for MD and FA was computed (bin widths: MD 0.000004 mm^2/s, FA 0.01; upper limit: MD 0.004 mm^2/s, FA 1.0) for each DTI voxel identified by the WM mask image. To correct for differences in brain volume, each histogram was normalized by the total number of voxels present with the WM mask. For each MD histogram, the normalized frequency of pixels at the peak height (MD NPH) was obtained and used for further analysis as these were deemed to be the most stable measurements. These parameters have been cross-sectionally shown to have associations with cognition in a SVD population.[8]

Estimation of Longitudinal WMH Volume Change

WMH masks were created using the above technique from the FLAIR images. The volumes at each time point were calculated in individual subject space by summing the binarized corrected segmentations. Three measurements were calculated, total brain volume (Grey Matter (GM)+White Matter (WM)), WMH volume, and WMH volume as a percentage of total brain volume (WMH/total brain volume × 100). The latter value was used for further analysis as brain volume changes with time and a reduced WM volume, for example, may underestimate growth of the WMHs. In a further step, we performed a scan–rescan test on two patients to determine the effect of changing head position on WMH volumes and found these to be identical using the automated technique.

Identification of Lacunes

Lacunes were manually identified in native subject space at each time point by a consultant neuroradiologist, using T1-weighted, T2-weighted, and FLAIR images. A lacune was defined as a CSF filled cavity, 3 to 15 mm in diameter with a surrounding rim of FLAIR hyperintensity.[16] Cavity size thresholds were chosen as lesions that are <3 mm in diameter are more likely to be perivascular spaces than lacunes and cavities >15 mm

are less likely to reflect an underlying SVD etiology.[16] For the longitudinal analysis, the images were first registered to a group average template. Our neuroradiologist used these registered images to identify new lacunes.

Neuropsychologic Assessment

Cognitive assessment was carried out annually using well-established standardized tests to include measures sensitive to the pattern of cognitive impairment associated with SVD. Premorbid Intelligent Quotient was estimated using the National Adult Reading Test-Restandardized.[23] Tasks were grouped into broad cognitive functions: executive function and processing speed.

Executive function: Trail making test, (Matrix B) (Reitan 1996) Modified Wisconsin Card Sorting Test,[24] and Phonemic Fluency (FAS).[25]

Processing speed: Wechsler Adult Intelligence Scale-III (Wechsler, 1997a) Digit symbol substitution,[26] Speed of Information Processing Task,[27] and Grooved Pegboard Task.[28]

Task performance was age scaled using manually published normative data, transformed into z-scores, and aggregated to construct the cognitive indices by averaging across the component test measures for each subject. For further details on the cognitive assessment, please refer to Lawrence et al.[8]

Statistical Analysis

Because of the inherited hierarchical nature of the data imposed by multiple measurements per patient, we use Linear Mixed Effects (LME) modeling to estimate change over the follow-up period in our MRI markers and cognitive measures.[29] Following graphical inspection of the raw data, it was reasonable to assume a multivariate normal distribution as well as linearity of change with time. Specifically, we used a random intercept and random slopes model, which permits the estimation of an average slope across the whole cohort while allowing for inter-individual variability.[29] The average slope represents the average annual change in a measurement and can be used for estimation of sample sizes. A Bayesian approach for statistical inference has been used and the parameter estimates are summarized as the means of their distribution and the uncertainty as their 95% credible interval (95% CI). This inferential approach accommodates for patient drop-out during the follow-up with the assumption that unobserved measurements are missing at random (MAR). The WinBUGS package (version 1.4.3, MRC Biostatistics Unit, Cambridge Institute of Public Health, Cambridge Biomedical Campus, Cambridge, UK) was used for estimation.

Sample size estimations for brain volume, WMH volume, DTI measures, and cognition were carried out using the longpower statistical package in R version 3.02 (http://www.R-project.org/) by using the estimated values for the intercept and slope variance and the residual variance.[30,31] The input for these calculations is presented in Table 3.

For lacunes, we defined the event as the occurrence of more lacunes than the number at baseline and the data have been modelled as for a binary outcome. We estimated the odds ratios of the event in association with the baseline number of lacunes and hence the predicted probability of developing more lacunes than the number found at the baseline. The proportion of developing new lacunes at some point during the follow-up period was then determined, and sample sizes were calculated for a percentage reduction in the average probability of having one or more new lacunes during the 3-year follow-up period.

In clinical trials, MRI data are often only available from two time points, baseline and final follow-up. We therefore used LME models to estimate average slopes using data only from two time points (baseline and 2 years, baseline and 3 years). All patients were included in the model as described above. Sample size estimations using data from two time points was compared with sample sizes using data from all four time points.

We also investigated the effect of missing data on the original estimates provided under the MAR assumption by using a simultaneous model having as components the random effect model as originally considered and a random effects logistic model with 'missingness' as a response (a binary longitudinal response where 1 designates the absence of a measurement for any reason and 0 its presence). This approach allowed us to investigate the relationship between data missing at follow-up and baseline measurements as well as the slope.

Results

Patient Demographics

Of the 121 participants at baseline in the SCANS study, 99 subjects returned at one or more time points: 98 at year 1, 77 at year 2, and 70 at year 3 (Table 2). One subject attended the baseline and missed the year 1 follow-up but attended all subsequent sessions. Four subjects missed the year 2 follow-up but subsequently attended at year 3. Of the 22 subjects who did not have follow-up MRI, 7 subjects died, 10 subjects formally withdrew from the study, 4 were lost to follow-up, and 1 declined further MRI. Demographics at baseline are shown in Table 2. Patients who left the study (for any reason) had a higher lacune load ($P < 0.013$), a higher mean Rankin disability score ($P < 0.026$), and

Table 2. Patient demographics at baseline.

Demographics and risk factors	N = 120
Mean age (s.d.), years	70 (9.8)
Mini Mental test score (mean (s.d.))	27.6 (2.7)
Sex	
Female	42 (35.0%)
Male	78 (65.0%)
Hypertension	
No	9 (7.5%)
Yes	111 (92.5%)
Hypercholesterolemia	
No	17 (14.2%)
Yes	103 (85.8%)
Smoker	
Never	55 (45.8%)
Current	23 (19.2%)
Ex-smoker	42 (35.0%)
Treated diabetes mellitus	
Yes	22 (18.3%)
No	96 (80.0%)
Diet control	2 (1.6%)
Rankin disability score	
0	38
1	48
2	15
3	16
4	3
MRI measures at baseline (per subject)	
Average (95% CI) number of lacunes	4 (4, 5.1)
Average (95% CI) WMH load as the pecentage of total normalized brain volume	3.72 (3.16, 4.28)
Average (95% CI) normalized brain volume in ml	1295 (1279.00, 1312.00)

Abbreviations: CI, credibility interval; WMH, white matter hyperintensity. Hypertension was defined as systolic blood pressure >140 mmHg or diastolic >90 mmHg or those on antihypertensive treatment. Hypercholesterolemia was defined as a serum total cholesterol >5.2 mmol/l or treatment with a statin.

a lower mean Mini Mental Test Examination ($P < 0.004$) score at baseline when compared with patients who attended all time points. There were, however, no significant differences in baseline brain volume, diffusion parameters, or other demographic characteristics.

Brain Volume

Over the course of the observational period, there was strong evidence of a decrease in brain volume. The average (95% credibility interval) annual brain volume change was equal to −8.86 m/s (−10.67, −7.05 or −0.69% of baseline per year; Table 3; Figure 1).

WMH Volume

Larger WMH volume at baseline was associated with a faster rate of growth. Over the course of the observational period, there was a strong evidence of growth in WMH volume. The average (95% credibility interval) annual change in WMH volume was equal to 0.80% (0.67, 0.95; Table 3; Figure 1).

Table 3. Change in MRI and cognitive indices estimated using a random intercepts and random slopes model fit to the data.

Parameter	Mean intercept	Mean slope	Slope variance	Residual variance
MRI measures				
Brain volume (ml)	1295.00 (1279.00, 1312.00)	−8.83 (−10.61, 7.01)	28.15 (5.63, 54.68)	182.10 (145.90, 227.40)
White matter hyperintensity (%)	3.72 (3.16, 4.28)	0.80 (0.67, 0.95)	0.43 (0.27, 0.65)	0.12 (0.09, 0.15)
MD peak height	0.015 (0.015, 0.016)	-3.87×10^{-4} $(-4.51 \times 10^{-4},$ $-3.24 \times 10^{-4})$	1.55×10^{-4} $(8.59 \times 10^{-5},$ $2.81 \times 10^{-4})$	5.16×10^{-4} $(4.62 \times 10^{-4},$ $5.77 \times 10^{-4})$
Cognitive indices				
Executive function	−0.92 (−1.12, −0.73)	0.0362 (−0.083, 0.010)	0.020 (0.053, 0.0384)	0.13 (0.10, 0.16)
Processing speed	−0.81 (−0.98, −0.64)	0.014 (−0.053, 0.024)	0.013 (0.022, 0.026)	0.092 (0.073, 0.126)

Abbreviations: MD, mean diffusivity; MRI, magnetic resonance imaging. 95% credible intervals are shown in brackets.
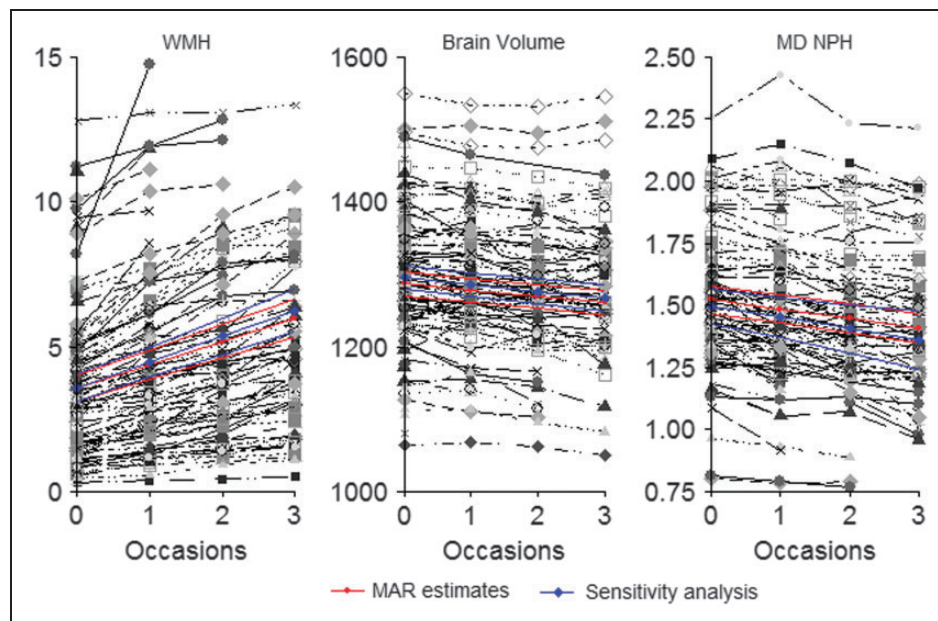
Figure 1. Plots showing individual trajectories in magnetic resonance imaging (MRI) markers showing a decrease in mean diffusivity (MD) normalized peak height (MD NPH), a decrease in brain volume and an increase in white matter hyperintensity volume (WMH) over the 3-year follow-up period. Time points are shown on the x axis. The average slope is shown in red with credibility intervals. There is only a minimal change in slope when missing data are accounted for in simultaneous models (shown in blue). MAR, missing at random.

Diffusion Imaging

MD NPH: There was strong evidence of a decrease in MD NPH over the course of the observational period. The average (95% credibility interval) annual change was equal to -3.87×10^{-4} (-4.51×10^{-4}, -3.24×10^{-4}) (or -2.54% with respect to baseline per year) (Table 3; Figure 1).

Lacunes

There were 98, 77, and 70 subjects with suitable T1-weighted images at time points 1, 2, and 3, respectively. Sixteen out of the 98 subjects had one or more new lacunes after 1 year of follow-up; 17 out of the 77 subjects had one or more new lacunes after 2 years of follow-up. Twenty out of the 70 subjects had one or more new lacunes after 3 years of follow-up. The number of lacunes at baseline was positively associated with a higher probability of new lacunes at follow-up. After modeling the baseline data using a Poisson distribution, the mean (95% credibility interval) number of baseline lacunes was estimated to be 4 (4.2, 5.1). However, 50% of subjects had 2 lacunes at baseline. The predicted probability of developing >4 lacunes over the 3-year follow-up period was estimated to be 0.21 (0.13, 0.31).

Cognition

Executive function: There was no evidence of change in executive function (Z-score) over the course of the observational period. The average (95% credibility interval) annual change was equal to -3.62×10^{-2} (-8.33×10^{-2}, 1.07×10^{-2}; Table 3).

Processing speed: There was no evidence of change in processing speed (Z-score) over the course of the observational period. The average (95% credibility interval) annual change was equal to -1.44×10^{-2} (-5.35×10^{-2}, 2.47×10^{-2}; Table 3).

Sample Size Calculations

Sample sizes are calculated for a hypothetical trial of 3-year trial duration with measurements taken every year and evenly in time. A balanced design is assumed. For brain volume, WMH growth, and MD NPH, the minimum sample size per arm required to detect a 30, 25, 20, and 15% treatment effect at 80% power was estimated. The results are shown in Table 4.

For lacunes, the sample size was calculated for a percentage reduction in the average probability of having >4 lacunes at follow-up. For a 3-year trial duration, the minimum sample size required to detect a 30, 25, 20, and 50% treatment effect at 80% power was 572; 842; 1,345; and 2,442; respectively (Table 4).

Table 4. The predicted minimum sample size per arm (for MRI and cognitive measures) (power = 0.0.8 and type 1 error = 0.05) for a hypothetical clinical trial of 3-year duration assuming a balanced design with measurements taken every year evenly in time to test hypothetical treatment effects of 30, 25, 20, 15, and 10% in the intervention group.

Parameter	Sample size (per arm) to test treatment effects of:			
	30%	25%	20%	15%
MRI measures				
WMH volume	124	178	279	496
Brain Volume	145	208	325	578
Lacunes	572	842	1,345	2,442
MD normalized peak height	128	185	289	513
Cognitive indices				
Executive function	6,135	8,834	13,803	24,539
Processing speed	26,369	37,972	59,331	105,478

Abbreviations: MD, mean diffusivity; MRI, magnetic resonance imaging; WMH, white matter hyperintensity.

Sample size estimates in the hypothetical scenario that only two MRI scans were performed during the trial are shown in Table 5. Estimates are given for a trial with a 2- and 3-year follow-up. Estimates were not greatly different from those using the LME modeling on all time points, particularly when 3-year follow-up data were used.

Sensitivity Analysis

The simultaneous models, which take 'missingness' into account, did not reveal dramatic changes in the average rate of change for the MRI measures. The 95% credibility intervals (for both intercepts and slopes) overlap to a large extent with the original MAR model (Supplementary Table S1; Figure 1). The simultaneous model for MD NPH showed the largest departure from its original MAR slope estimate of -0.039 (-0.045, -0.032) to -0.045 (-0.07, -0.018) (Supplementary Table S1, Figure 1).

The analysis showed that older patients were more likely to drop out of the study. There was no strong evidence that baseline WMH volume, brain volume or diffusion parameters were predictive of dropout however we found that patients with a faster increase in WMH volume were more likely to drop out at subsequent time points (Supplementary Table S1).

Baseline WMH volume was positively correlated with the rate of WMH growth. Larger brain volumes at baseline were also correlated with a faster rate of atrophy. There was, however, no relationship between baseline diffusion parameters and their rate of change (Supplementary Table S1).

Table 5. Sample size estimates for MRI markers using all available data (four time points) compared with sample size estimates using data only from two time points.

Variable	Percentage difference in the slope	Minimum sample size per arm (using all four time points)	Minimum sample size per arm (using only baseline and 3-year data)	Minimum sample size per arm (using only baseline and 2-year data)
WMH volume	(30%)	124	61	72
	(25%)	178	88	104
	(20%)	279	137	162
	(15%)	496	244	287
MD NPH	(30%)	128	138	261
	(25%)	185	199	375
	(20%)	289	311	587
	(15%)	513	553	1,044
Brain volume	(30%)	145	106	97
	(25%)	208	152	140
	(20%)	325	238	219
	(15%)	578	423	388

Abbreviations: MD NPH, mean diffusivity normalized peak height; MRI, magnetic resonance imaging; WMH, white matter hyperintensity. The final column shows sample size estimates if patients were followed up for only 2 years.

Discussion

This prospective longitudinal study in symptomatic SVD has shown that change in MRI markers is detectable over a 3-year follow-up period. In contrast over the same time period there was no evidence of change detectable in cognitive measures. The effect of age on cognitive and MRI measures was deemed negligible. It follows that sample size estimates of the number of patients needed to demonstrate a treatment effect in an interventional trial are much smaller when using MRI markers as an end point.

The lack of cognitive change is consistent with the recent results of the SPS3 trial in approximately 3,000 patients with MRI confirmed lacunar stroke.[32] Over a median follow-up of 3 years, there was very little cognitive change. Why cognitive change is so difficult to detect in such a cohort, when it is easily detectable in a high proportion of individuals in cross sectional studies[8] is unclear. Possible factors may include measurement error, practice effects, and importantly rapidly declining patients dropping out of the study which there was evidence of in our SCANS cohort. In addition, rates may differ in different patient cohorts with individuals in the more advanced stages of cognitive impairment tending to decline faster. However, whatever the reason, the difficulty detecting cognitive change highlights the need for more sensitive ways to detect change and allow screening of novel therapies in smaller number and short follow-up periods.

A useful surrogate marker needs to be both sensitive to change and to correlate with meaningful clinical outcomes. In this study, WMH volume and the diffusion parameter, MD NPH, were most sensitive to change and therefore had the smallest sample size estimates. In contrast to a previous smaller study with shorter follow-up,[12] DTI markers would require slightly larger sample sizes than WMH, but further data are required on the relative sensitivity of the two measures to white matter damage.

The available evidence indicates that the MRI markers investigated in this study do correlate with clinical outcomes as outlined in the introduction. Although WMH volume has been shown to be associated with cognitive impairment, these associations have been inconsistent and weak,[5,33] particularly in patients with more severe disease and confluent WMHs at study entry.[12] This was also the case in the baseline data from the SCANS cohort used in this study.[8] Brain volume and DTI parameters show stronger associations with cognitive impairment,[8,34] suggesting they may represent clinically meaningful surrogate markers. Deciding which markers are most useful can only be definitively evaluated by including them in clinical trials and determining which correlate most closely with treatment effects measured using clinical end points. A combination of these MRI measures (i.e., a composite score) may offer a more sensitive approach; however, whether such a parameter is able to predict change in a clinical outcome measure longitudinally is

yet to be determined. In addition, the relative sensitivity of clinical end points such as progression to disability and dementia will need to be assessed in larger cohorts with longer follow-up.

Because of the constraints of longitudinal follow-up and the cost of MRI scanning in a clinical trial, it is often not practical to collect MRI data for >2 time points. We therefore performed additional analyses to determine whether using MRI data from only two time points affected the sample size estimates. We found that sample size estimates for WMH appear slightly smaller when only data from baseline and 3 year follow-up are considered but remain largely similar for brain volume and MD NPH. This suggests that this approach is feasible and has similar power for clinical trials; however, it is important to be aware that these two time point models are more susceptible to noise and data MAR. In studies using MRI data from only two time points, the interval must always be long enough for sufficient change to accumulate and be detectable above the noise that is inevitable whenever measurements are made. This is evident from the higher estimated sample sizes with 2 vs. 3 year follow-up.

It is important to carefully define and identify suitable cases for clinical trials in SVD. The heterogeneous nature of SVD is a major hindrance to this. The cerebral small vessels can be affected by a wide range of pathologies, including cerebral amyloid angiopathy, immunologically mediated diseases, and genetic SVDs, such as CADASIL.[1] The most common type, however, is age-related and vascular risk factor-related SVD. There is increasing evidence to suggest that even this type of SVD may have multiple pathologies. It has been suggested there are at least two underlying arterial pathologies: microatheroma primarily causing larger, usually isolated lacunar infarcts and a diffuse arteriopathy characterized by arteriolosclerosis, lipohyalinosis, or fibrinoid necrosis (depending on severity of the abnormality), which is thought largely to result from hypertension and is associated with multiple smaller lacunar infarcts and leukoaraiosis.[35,36] This distinction is supported by pathologic, radiologic, and risk factor data.[37] In view of these possible differences and to study a homogenous group of patients, we recruited only patients with symptomatic lacunar infarction confirmed on MRI as well as confluent leukoaraiosis.

CMBs are a feature of SVD that are thought to vary in their distribution depending on the underlying disease process,[16] although their prognostic significance remains uncertain. We have analyzed CMBs in the baseline data and shown a weak correlation with cognition,[38] but longitudinal data on CMBs were not available for this analysis. Therefore, we were unable to determine sample size measures based on this imaging parameter.

A major limitation of this study was that we had a relatively high dropout rate, although this is consistent with previous longitudinal studies in ageing.[39] Patients who did not complete follow-up tended to be older and more disabled. This may have led to an underestimation of the rate of change in MRI markers and cognition. Optimal trial design has to take careful consideration of issues such as the dropout rate and increase sample sizes accordingly. Errors in image processing and registration are inevitable when attempting to quantify MRI markers in a standardized manner. This will increase noise in the data and reduce sensitivity to change. For example, the reported error of SIENA is 0.15%. It is therefore unreliable when the mean difference between groups is <0.2%. In addition, measurement error may be higher when these markers are used in a multicenter trial with imaging on different scanners. Because of the heterogeneous nature of the disease, it is possible that in some patients MRI markers have plateaued to maximal impairment or have not yet changed appreciably (ceiling and floor effects, respectively). Studies with longer follow-up may be capable of identifying these patients and capture potential non-linear aspects of change. Finally, these data were acquired on a 1.5-Tesla MRI scanner. The quality and precision of data might be improved by acquiring data using sequences with isotropic voxel dimensions, at higher resolutions, and MRI field strengths.[40]

In conclusion, MRI is promising as a potential surrogate marker in SVD. It is much more sensitive to change than cognitive measures over a relatively short time period. Optimizing and validating these markers could significantly reduce the size, duration, and cost of clinical trials. Further studies evaluating these markers and particularly relating them to clinical and cognitive outcomes in longitudinal studies are required, and this would ideally be done within the context of a randomized clinical trial.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Authors' contributions

Philip Benjamin—Conception, acquisition of data, analysis and interpretation of data, drafting the article. Eva Zeestraten—Analysis and interpretation of data, drafting the article. Christian Lambert—Analysis and interpretation of data. Irina Chis Ster—Statistical analysis. Owen A Williams—Analysis and interpretation of data. Andrew J Lawrence—Acquisition of data, analysis and interpretation of data. Bhavini Patel—Acquisition of data, analysis and interpretation of data. Andrew D MacKinnon—Analysis and interpretation of data. Thomas R Barrick—Supervisor, critical revision of article, final approval. Hugh S Markus—Supervisor, conception, critical revision of article, final approval.

Supplementary Material

Supplementary material for this paper can be found at http://jcbfm.sagepub.com/content/by/supplemental-data

References

1. Pantoni L. Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. *Lancet Neurol* 2010; 9: 689–701.
2. Patel B and Markus HS. Magnetic resonance imaging in cerebral small vessel disease and its use as a surrogate disease marker. *Int J Stroke* 2011; 6: 47–59.
3. Van Norden AGW, van Dijk EJ, de Laat KF, Scheltens P, OldeRikkert MGM and de Leeuw FE. Dementia: Alzheimer pathology and vascular factors: from mutually exclusive to interaction. *Biochim Biophys Acta* 2012; 1822: 340–349.
4. Schmidt R, Berghold A, Jokinen H, Gouw AA, van Flier WM, der, Barkhof F, et al. White matter lesion progression in LADIS frequency, clinical effects, and sample size calculations. *Stroke* 2012; 43: 2643–2647.
5. Holtmannspötter M, Peters N, Opherk C, Martin D, Herzog J, Brückmann H, et al. Diffusion magnetic resonance histograms as a surrogate marker and predictor of disease progression in CADASIL: a two-year follow-up study. *Stroke J Cereb Circ* 2005; 36: 2559–2565.
6. Lawrence AJ, Chung AW, Morris RG, Markus HS and Barrick TR. Structural network efficiency is associated with cognitive impairment in small-vessel disease. *Neurology* 2014; 83: 304–311.
7. Fleming TR and DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med* 1996; 125: 605–613.
8. Lawrence AJ, Patel B, Morris RG, MacKinnon AD, Rich PM, Barrick TR, et al. Mechanisms of cognitive impairment in cerebral small vessel disease: multimodal MRI results from the St George's Cognition and Neuroimaging in Stroke (SCANS) Study. *PLoS One* 2013; 8: e61014.
9. Arani Nitkunan SL. Brain atrophy and cerebral small vessel disease: a prospective follow-up study. *Stroke J Cereb Circ* 2011; 42: 133–138.
10. Jokinen H, Lipsanen J, Schmidt R, Fazekas F, Gouw AA, van der Flier WM, et al. Brain atrophy accelerates cognitive decline in cerebral small vessel disease: the LADIS study. *Neurology* 2012; 78: 1785–1792.
11. Peters N, Holtmannspötter M, Opherk C, Gschwendtner A, Herzog J, Sämann P, et al. Brain volume changes in CADASIL: a serial MRI study in pure subcortical ischemic vascular disease. *Neurology* 2006; 66: 1517–1522.
12. Nitkunan A, Barrick TR, Charlton RA, Clark CA and Markus HS. Multimodal MRI in cerebral small vessel disease its relationship with cognition and sensitivity to change over time. *Stroke* 2008; 39: 1999–2005.
13. Van Norden AGW, de Laat KF, van Dijk EJ, van Uden IWM, van Oudheusden LJB, Gons RAR, et al. Diffusion tensor imaging and cognition in cerebral small vessel disease: the RUN DMC study. *Biochim Biophys Acta* 2012; 1822: 401–407.
14. Benjamin P, Lawrence AJ, Lambert C, Patel B, Chung AW, MacKinnon AD, et al. Strategic lacunes and their relationship to cognitive impairment in cerebral small vessel disease. *Neuroimage Clin* 2014; 4: 828–837.
15. Jokinen H, Gouw AA, Madureira S, Ylikoski R, van Straaten ECW, van der Flier WM, et al. Incident lacunes influence cognitive decline: the LADIS study. *Neurology* 2011; 76: 1872–1878.
16. Wardlaw JM, Smith EE, Biessels GJ, Cordonnier C, Fazekas F, Frayne R, et al. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *Lancet Neurol* 2013; 12: 822–838.
17. Fazekas F, Kleinert R, Offenbacher H, Schmidt R, Kleinert G, Payer F, et al. Pathologic correlates of incidental MRI white matter signal hyperintensities. *Neurology* 1993; 43: 1683–1689.
18. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 2004; 23: S208–S219.
19. Altmann DR, Jasperse B, Barkhof F, Beckmann K, Filippi M, Kappos LD, et al. Sample sizes for brain atrophy outcomes in trials for secondary progressive multiple sclerosis. *Neurology* 2009; 72: 595–601.
20. Anderson VM, Schott JM, Bartlett JW, Leung KK, Miller DH and Fox NC. Gray matter atrophy rate as a marker of disease progression in AD. *Neurobiol Aging* 2012; 33: 1194–1202.
21. Ashburner J and Friston KJ. Diffeomorphic registration using geodesic shooting and Gauss-Newton optimisation. *Neuroimage* 2011; 55: 954–967.
22. Lambert C, Lutti A, Helms G, Frackowiak R and Ashburner J. Multiparametric brainstem segmentation using a modified multivariate mixture of Gaussians. *Neuroimage Clin* 2013; 2: 684–694.
23. Nagahama Y, Okina T, Suzuki N, Matsuzaki S, Yamauchi H, Nabatame H, et al. Factor structure of a modified version of the Wisconsin Card Sorting Test: an analysis of executive deficit in Alzheimer's disease and

mild cognitive impairment. *Dement Geriatr Cogn Disord* 2003; 16: 103–112.

24. Delis DC, Kaplan E and Kramer JH. *Delis-Kaplan executive function system*. San Antonio, TX: Psychological Corporation, 2001.

25. Wechsler D. *Wais-III, Wechsler Adult Intelligence Scale, Third Edition: WMS-III, Wechsler Memory Scale*, 3rd ed. Technical Manual. Psychological Corporation, San Antonio, TX, 1997.

26. Coughlan A, Oddy M and Crawford J. *The BIRT memory and information processing battery (B-MIPB)*. Wakefield, UK: The Brain Injury Rehabilitation Trust (BIRT), 2009.

27. Matthews C and Klove H. *Instruction manual for the adult neuropsychological test battery*. Madison, WI: University of Wisconsin Medical School, 1964.

28. Diggle PJ, Heagerty P, Liang K-Y and Zeger SL. *Analysis of longitudinal data*. Oxford University Press, 2002.

29. Liu G and Liang K-Y. Sample size calculations for studies with correlated observations. *Biometrics* 1997; 53: 937–947.

30. Lu K, Luo X and Chen P-Y. Sample size estimation for repeated measures analysis in randomized clinical trials with missing data. *Int J Biostat* 2008; 4, article 9: 1–16.

31. Pearce LA, McClure LA, Anderson DC, Jacova C, Sharma M, Hart RG, et al. Effects of long-term blood pressure lowering and dual antiplatelet treatment on cognitive function in patients with recent lacunar stroke: a secondary analysis from the SPS3 randomised trial. *Lancet Neurol* 2014; 13: 1177–1185.

32. Liem MK, Lesnik Oberstein SJ, Haan J, van der Neut IL, Ferrari MD, van Buchem MA et al. MRI correlates of cognitive decline in CADASIL: a 7-year follow-up study. *Neurology* 2009; 72: 143–148.

33. Charlton RA, Barrick TR, McIntyre DJ, Shen Y, O'Sullivan M, Howe FA, et al. White matter damage on diffusion tensor imaging correlates with age-related cognitive decline. *Neurology* 2006; 66: 217–222.

34. Fisher CM. The arterial lesions underlying lacunes. *Acta Neuropathol (Berl)* 1968; 12: 1–15.

35. Boiten J, Lodder J and Kessels F. Two clinically distinct lacunar infarct entities? A hypothesis. *Stroke* 1993; 24: 652–656.

36. Khan U, Porteous L, Hassan A and Markus HS. Risk factor profile of cerebral small vessel disease and its subtypes. *J Neurol Neurosurg Psychiatry* 2006; 78: 702–706.

37. Patel B, Lawrence AJ, Chung AW, Rich P, MacKinnon AD, Morris RG, et al. Cerebral microbleeds and cognition in patients with symptomatic small vessel disease. *Stroke* 2013; 44: 356–361.

38. Glymour MM, Chêne G, Tzourio C and Dufouil C. Brain MRI markers and dropout in a longitudinal study of cognitive aging. *Neurology* 2012; 79: 1340–1348.

39. Vrenken H, Jenkinson M, Horsfield MA, Battaglini M, van Schijndel RA, Rostrup E, et al. Recommendations to improve imaging and analysis of brain lesion load and atrophy in longitudinal studies of multiple sclerosis. *J Neurol* 2013; 260: 2458–2471.

40. Garde E, Lykke Mortensen E, Rostrup E and Paulson OB. Decline in intelligence is associated with progression in white matter hyperintensity volume. *J Neurol Neurosurg Psychiatry* 2005; 76: 1289–1291.

41. Van den Heuvel DMJ, ten Dam VH, de Craen AJM, Admiraal-Behloul F, Olofsen H, Bollen ELEM, et al. Increase in periventricular white matter hyperintensities parallels decline in mental processing speed in a nondemented elderly population. *J Neurol Neurosurg Psychiatry* 2006; 77: 149–153.

42. Schmidt R, Ropele S, Enzinger C, Petrovic K, Smith S, Schmidt H, et al. White matter lesion progression, brain atrophy, and cognitive decline: the Austrian stroke prevention study. *Ann Neurol* 2005; 58: 610–616.