Automated White Matter Total Lesion Volume Segmentation in Diabetes

J.A. Maldjian, C.T. Whitlow, B.N. Saha, G. Kota, C. Vandergriff, E.M. Davenport, J. Divers, B.I. Freedman, and D.W. Bowden

ABSTRACT

BACKGROUND AND PURPOSE: WM lesion segmentation is often performed with the use of subjective rating scales because manual methods are laborious and tedious; however, automated methods are now available. We compared the performance of total lesion volume grading computed by use of an automated WM lesion segmentation algorithm with that of subjective rating scales and expert manual segmentation in a cohort of subjects with type 2 diabetes.

MATERIALS AND METHODS: Structural T1 and FLAIR MR imaging data from 50 subjects with diabetes (age, 67.7 ± 7.2 years) and 50 nondiabetic sibling pairs (age, 67.5 ± 9.4 years) were evaluated in an institutional review board–approved study. WM lesion segmentation maps and total lesion volume were generated for each subject by means of the Statistical Parametric Mapping (SPM8) Lesion Segmentation Toolbox. Subjective WM lesion grade was determined by means of a 0–9 rating scale by 2 readers. Ground-truth total lesion volume was determined by means of manual segmentation by experienced readers. Correlation analyses compared manual segmentation total lesion volume with automated and subjective evaluation methods.

RESULTS: Correlation between average lesion segmentation and ground-truth total lesion volume was 0.84. Maximum correlation between the Lesion Segmentation Toolbox and ground-truth total lesion volume ($\rho = 0.87$) occurred at the segmentation threshold of $k = 0.25$, whereas maximum correlation between subjective lesion segmentation and the Lesion Segmentation Toolbox ($\rho = 0.73$) occurred at $k = 0.15$. The difference between the 2 correlation estimates with ground-truth was not statistically significant. The lower segmentation threshold (0.15 versus 0.25) suggests that subjective raters overestimate WM lesion burden.

CONCLUSIONS: We validate the Lesion Segmentation Toolbox for determining total lesion volume in diabetes-enriched populations and compare it with a common subjective WM lesion rating scale. The Lesion Segmentation Toolbox is a readily available substitute for subjective WM lesion scoring in studies of diabetes and other populations with changes of leukoaraiosis.

ABBREVIATIONS: DM = type 2 diabetes mellitus; LST = Lesion Segmentation Toolbox; SPM8 = Statistical Parametric Mapping; TLV = total lesion volume; GT = ground-truth

Leukoariaosis is a common WM pathologic lesion in older adults, characterized histologically by demyelination, loss of oligodendrocytes, and vacuolization resulting from small-vessel

ischemia of the WM.[1] On brain MR imaging, these lesions are commonly termed WM hyperintensities and appear as regions of increased signal on T2-weighted and FLAIR sequences. Increases in WM disease burden have been associated with risk factors, such as hypertension, type 2 diabetes mellitus (DM), and tobacco use.[2] Quantifying WM disease burden in the brain is important because it is an accurate and sensitive predictor of future stroke, dementia, and cognitive decline.[3-6]

Clinical evaluation of WM disease has been limited to subjective interpretation of disease burden, with typical modifiers including "few scattered lesions" and "mild," "moderate," or "severe" applied in radiologic reporting. This is because of the onerous and time-consuming task of manual delineation of WM lesion burden and the lack of robust automated tools for quantitative WM lesion grading. A semi-quantitative visual rating scheme was developed for use with large epidemiologic studies

involving brain MR imaging. This WM hyperintensity grading scale is commonly used for research studies and is based on visual assessment by an experienced reader with the use of a semi-quantitative 10-point (0–9) scale with predefined methodology.[7,8]

A variety of automated methods for WM lesion quantification, involving combinations of thresholding, segmentation, prior information, lesion growing algorithms, and, most recently, machine learning algorithms, have been used in research studies of multiple sclerosis. It is beyond the scope of this work to describe all the developments in WM lesion segmentation. Rather, we focus on a recently described software tool for automated WM lesion segmentation, the Lesion Segmentation Toolbox (LST); it was developed for use in the Statistical Parametric Mapping (SPM8) environment, which is familiar to most neuroimaging researchers, is freely distributed open source code written in Matlab (MathWorks, Natick, Massachusetts), and is easily implemented/fully automated. Additionally, the LST uses widely available structural T1-weighted and FLAIR images for performing WM lesion segmentation. The LST was developed for use in multiple sclerosis and originally evaluated in a group of 52 subjects with multiple sclerosis and 18 control subjects, achieving excellent agreement with manual tracing (R^2 values of 0.93).[9] A critical user-determined parameter in the LST procedure is the k-threshold, which was determined to be 0.3 in a multiple sclerosis population.

The purpose of our study is to compare the performance of total lesion volume (TLV) grading computed by means of the LST automated WM lesion segmentation algorithm to expert manual segmentation in a cohort of subjects with DM and to determine the optimum k-threshold in this population. A secondary objective is to compare the LST TLV with semi-quantitative subjective rating scales. Our hypothesis is that TLV computed with the LST will perform at least as well as subjective rating scales when compared with ground-truth (GT) TLV in DM. This report answers the important question of whether an automated toolbox developed for use in multiple sclerosis can be used reliably for grading WM lesion burden in populations with a different pathophysiologic mechanism for development of WM disease.

MATERIALS AND METHODS
Subjects
The Diabetes Heart Study is a genetic and epidemiologic study of 1443 European American and African American participants from 564 families with multiple cases of DM.[10,11] The Diabetes Heart Study–Mind is an extension of the Diabetes Heart Study family of studies and examines the genetic and brain imaging contributors to cognitive changes associated with DM. The study includes diabetes- and nondiabetes-affected siblings. All subjects provided written informed consent, and study protocols were approved by the blinded institutional review board. MR imaging studies from 100 subjects were randomly selected from the Diabetes Heart Study-Mind. These included 50 subjects with DM (27 women, 23 men), 52% smokers, mean ± standard deviation (SD) age of 67.7 ± 7.2 (age range, 52–84 years), body mass index of 32.3 ± 7.1, and hemoglobin A1C of 7.6 ± 1.48; and 50 siblings without DM (35 women, 15 men), 46% smokers, mean ± SD age of 67.5 ± 9.4 (age range, 43–89 years), body mass index of 28.5 ± 6.5, and hemoglobin A1C of 5.9 ± 0.31.

MR Imaging
Participants from the Diabetes Heart Study-Mind were scanned on a 1.5T scanner with twin-speed gradients, with the use of an 8-channel neurovascular head coil Twin Speed EXCITE; GE Healthcare, Milwaukee, Wisconsin. High-resolution T1 anatomic images were obtained by means of a 3D spoiled gradient-echo sequence (matrix, 256 × 256; field of view, 20 cm; section thickness, 1.5 mm with no gap; number of sections, 124; in-plane resolution, 0.781 × 0.781 mm) aligned parallel to the anterior/posterior commissures (anterior/posterior commisure line). FLAIR images were acquired in the axial plane for the purpose of evaluating WM hyperintensities (TR = 8002, TE = 108.5, TI = 2000, flip angle = 90, 24 cm FOV, matrix size = 256 × 256 [0.94 × 0.94 mm], 3-mm section thickness).

Semi-Quantitative WM Rating Scale
WM hyperintensity signal changes of each individual were assessed independently by 2 board-certified neuroradiologists by means of a semi-quantitative 10-point (0–9) scale with predefined methodology.[7,8] WM hyperintensity burden was estimated as the total extent of periventricular and subcortical white matter FLAIR signal hyperintensity that successively increases from no or barely detectable changes (grades 0 and 1, respectively) to almost all WM involved (grade 9). This scale has an inter-reader reliability agreement within 1 grade of 85.7%, with relaxed κ of 0.8, and intrareader reliability for agreement within 1 grade of 96.9%, with relaxed κ of 0.96.[8]

Image Preprocessing
The structural T1-weighted images were segmented into gray matter, WM, and CSF, normalized to Montreal Neurological Institute imaging space, and modulated with the Jacobian determinants of the normalization procedure to obtain tissue volume maps by use of the Dartel high-dimensional warping and the SPM8 (Wellcome Department of Imaging Neuroscience, London, UK)[12] new segment procedure, as implemented in the VBM8 toolbox (http://dbm.neuro.uni-jena.de/vbm.html). In addition to normalized images in Montreal Neurological Institute imaging space, the procedure outputs native space segmentations and a native space partial volume estimate label image of the most likely tissue class for each voxel. The quality of the segmentation and normalization for all subjects was confirmed by visual inspection.

WM Lesion Segmentation
WM lesion segmentation and TLV maps were generated by use of the LST[9] for SPM8 at 20 thresholds (*k*), ranging from 0–1 at 0.05 increments. The algorithm operates in native space and initially coregisters the FLAIR images to the space of the native T1. Each voxel in the T1 image is assigned to 1 of 3 classes (gray matter, WM, CSF) by use of the VBM8 toolbox as described above (partial volume estimate label map in native space). The FLAIR intensity distribution is calculated for each of the 3 classes to determine outliers, weighted according to the spatial probability of being WM, resulting in 3 classes of belief maps, and summed to generate a single belief map. A binarized version of the gray matter lesion map is used to seed a region growing algorithm

with the summed belief map as the target. The user-selected k-threshold is used as the cutoff to generate the initial gray matter seed binarized belief map. The algorithm outputs WM lesion segmentations for each k threshold, as well as a table of total lesion volume (20 TLV values corresponding to each of the 20 k-thresholds).

GT Segmentation

A multi-tiered approach was used to generate the reference GT lesion volume segmentation. All stages involved expert manual segmentation by board-certified neuroradiologists. Two raters (each with 1–2 years of neuroradiology experience) initially manually segmented all the white matter lesions independently by use of in-house software. This generated 2 independent lesion volume datasets. The 2 datasets were merged by use of a binary union procedure to generate a single dataset representing the combination of all lesions identified by both raters. Two highly experienced neuroradiologists (with >15 years and >3 years of experience, respectively) then together performed a consensus reading, reviewing all lesions in the combination dataset. The consensus reading session was performed by use of MRIcron (http://www.sph.sc.edu/comd/rorden/mricro.html)[13] with direct overlay onto the original FLAIR images and included the ability to manually add, remove, and edit the borders of all lesions. This final consensus lesion volume dataset served as the GT for the study.

Statistical Analysis

Correlation and regression analyses were performed for 3 main comparisons: subjective WM lesion scores versus GT, LST versus GT for each k-threshold, and subjective WM lesion scores versus LST for each k threshold. These analyses were repeated by use of log (TLV) to account for nonlinearity in the relationship between TLV and subjective scores. The optimum k-threshold was determined as that providing the maximum correlation value compared with manual segmentation. Between-group and inter-reader comparisons were also performed for the subjective WM scores and correlations with GT. Fisher r-to-z transformation[14] was used to test whether observed correlations were statistically different from zero. Steiger Z-test of correlated correlations[15] was used to determine if there was a significant difference in correlation values between subjective rating with GT and LST with GT. Bland-Altman plots were also computed between LST versus GT and LST versus subjective WM lesion scores to determine the mean difference between methods.

RESULTS

Subjective WM Lesion Scores

Mean WM score for the group was 2.0, with an SD of 1.5. The distribution of WM scores ranged from 0–7, with most values falling in the lower range of 0–3 (On-line Fig 1). Between-reader agreement was 88% within 1 grade, similar to that reported in the literature for this method.[8] There was no statistically significant difference in subjective WM scores between groups, with mean ± SD of 2.0 ± 1.3 for DM-affected and 2.0 ± 1.6 for non-DM groups.

Subjective WM Lesion Scores Versus GT

The Pearson correlation between average WM scores and GT was 0.84, with reader 1 having a correlation of 0.79 and reader 2 a correlation of 0.82. The correlation between average WM scores and GT was 0.82 and 0.85 for DM-affected and non-DM groups, respectively. With the use of the logarithm of GT, the correlations with average WM scores improved to 0.85 for the entire cohort (0.77 for reader 1 and 0.86 for reader 2) and 0.845 for the DM-affected group and 0.86 for the non-DM group. All reported correlation values were statistically significant ($P \leq$ .0001).

LST Versus GT

On-line Fig 2 is a graph of the LST k-threshold versus correlation with GT for the entire cohort. Correlation values ranged from 0.62–0.81 for the thresholds considered. Maximum correlation for the full sample was 0.87, corresponding to $k = 0.25$. For the non-DM group, maximum correlation was 0.94 observed at $k = 0.15$. For the DM-affected group, the maximum correlation was 0.83, corresponding to $k = 0.25$. The GT correlation performance of the LST was not significantly different from subjective WM scores (Steiger $Z = 0.8$, $P < .4$ for raw GT and $Z = 1.4$, $P < .15$ for the log-transformed data).

LST Versus Subjective WM Lesion Scores

Correlation values ranged from 0.59–0.65 for the threshold values considered. Maximum correlation was 0.74, which was observed when $k = 0.15$ in the full sample (On-line Fig 2). For the non-DM group, maximum correlation was 0.76, which was observed at the same threshold of $k = 0.15$. For the DM-affected group, maximum correlation was 0.74, corresponding to a k-threshold of 0.2. The threshold value of 0.25 provided correlations of 0.73, 0.74, and 0.74 for the full sample, the non-DM group, and the DM-affected group, respectively. Correlation values were similar with the use of log-transformed values.

GT Manual Segmentation

Paired Spearman correlation between manual segmentation raters, the final consensus GT volume, and the LST were all very high ($\rho = 0.91$ between first and second rater, $\rho = 0.96$ between first rater and GT, $\rho = 0.98$ between second rater and GT, $\rho = 0.92$ between first rater and LST, $\rho = 0.8$ between second rater and LST, and $\rho = 0.87$ between GT and LST), with $P < .001$ for all comparisons. Although the inter-rater correlations for manual segmentation were high, mean ± SD (median) of lesion volumes reported in units of milliliters were more widely spaced between raters: GT and LST with 2.43 ± 3.97 (0.86) for rater 1, 3.65 ± 5.69 (1.2) for rater 2, 4.47 ± 6.48 (1.83) for GT, and 2.45 ± 4.59 (0.38) for the LST, which suggests high inter-rater variability. The GT mean ± SD TLV for the full sample was 4.47 ± 6.48. The distribution of WM TLV was heavily skewed toward values <1 mL (On-line Fig 3), with a minor secondary peak at 10 mL. The mean ± SD TLV for the DM-affected group was 4.42 ± 4.62. Mean ± SD TLV for the non-DM group was 4.52 ± 7.13. There was no statistically significant difference in TLV between groups. On-line Fig 4 is a plot of GT TLV versus LST TLV at a k-threshold of 0.25, demonstrating a strong linear relationship ($R^2 = 0.76$).

On-line Fig 5 is a plot of GT TLV versus average WM scores demonstrating a nonlinear relationship for the subjective scale. To perform nonlinear fitting of these data, the WM scores were remapped from 0–9 to 1–10, and 2 data points with zero GT TLV were excluded. Logarithmic, polynomial, and power law relationships all provided better fits than a linear fit. The highest R^2 (0.72) was provided by a logarithmic fit. Plotting WM scores versus log(GT TLV) demonstrated a strong linear relationship ($R^2 = 0.73$).

DISCUSSION

The LST was originally developed and evaluated for multiple sclerosis. Here, we validate its use for TLV measurement in subjects with DM, relative to subjective rating scales. The pathophysiologic mechanisms leading to visible MR imaging WM changes in DM are very different from those in multiple sclerosis. In multiple sclerosis, the prototypic hallmark is focal demyelination with varying degrees of gliosis and inflammation.[16] The MR imaging FLAIR appearance is typically that of focal well-demarcated round or ovoid lesions. In contrast, WM lesions in the elderly, or leukoaraiosis, tend to be more diffuse, with pathophysiology related to endothelial dysfunction and development of small-vessel ischemia.[1,17-19] In this regard, diabetic populations provide an important validation of the LST methodology that can potentially be extended to other populations in which WM lesion burden relates to microvascular disease. Additionally, we examined a relatively large cohort of non-DM siblings, providing a validation in a normal elderly population.

LST Versus Subjective WM Lesion Scores

We demonstrate the LST to be comparable to subjective WM scores for determining severity of WM lesion load in a population with DM. A high degree of correlation was observed between TLV computed by use of the LST and GT manual segmentation. LST achieved a maximum correlation of 0.87, corresponding to a k-threshold of 0.25, and appeared robust in its segmentation performance, within a range of k-thresholds from 0.2–0.4, all demonstrating similar correlations. In comparison, the subjective WM scoring demonstrated slightly weaker correlation with GT, achieving a 0.84 correlation. This improved slightly to 0.85 by use of the log of the GT TLV. Additionally, the subjective WM scores demonstrated a nonlinear relationship to GT TLV. This is not surprising because the amount of visible WM disease increases substantially over the range of scores. That is, near total involvement of the brain WM for a grade of 9 is much greater in volume than 10 times the few lesions identified for a grade of 1. The use of a logarithmic transformation of the GT TLV provided a clear linear relationship to the subjective WM lesion scores. This finding has important implications for studies that use the subjective WM lesion rating scale. Studies that use the subjective WM lesion rating scale with standard statistical regression models violate assumptions of linearity and potentially affect the validity or significance of the results. The assumptions required for use of parametric testing (eg, Pearson correlation analysis) in this evaluation were fulfilled through log transform of the data. Alternatively, nonparametric testing can be used when these assumptions are not met. Repeating these analyses by use of the Spearman rank

correlation test and Kendall τ again demonstrated higher values for the LST than subjective WM scores, but the differences did not achieve significance. The Pearson correlation, however, provides a more complete assessment of the associations between variables when the underlying assumptions are attained.

The k-threshold corresponding to the maximum correlation between subjective scoring and LST was lower than that for LST and GT (0.15 versus 0.25). The lower k value in the LST corresponds to a more relaxed threshold for detection of WM lesions. This suggests that subjective ratings overestimate the true degree of WM disease. The degree of reader bias toward overestimation may be greater at the lower disease burden range, which was typical of our sample.

LST Versus GT

LST achieved a maximum R^2 of 0.69 at a k-threshold of 0.25 in the sample with DM. In contrast, the LST achieved a maximum R^2 of 0.94 and optimum k-threshold of 0.3 in the recent evaluation of multiple sclerosis.[9] The performance difference between DM and multiple sclerosis probably is an effect of disease severity. In the multiple sclerosis evaluation, there was greater disease burden, with TLV extending to >50 mL, compared with <35 mL for our DM population. More importantly, the performance in the multiple sclerosis population improved with increasing lesion volume, ranging from a mean Dice coefficient of 0.67 for lesion volumes <5 mL to 0.85 for lesions volume >15 mL. The multiple sclerosis evaluation did not appear to have a significant number of subjects with TLV <1 mL (if any), whereas for our population, most subjects had TLV <1 mL. Thus, the difference in performance for our group probably reflects disease severity, with the LST performing less optimally at low disease burdens, rather than a difference in lesion detectability between populations.

Automated Segmentation in Diabetes

Although automated WM lesion detection is an area of active investigation, there have been very few studies validating tools in a diabetic population. Jongen et al[20] described a k-nearest neighboring clustering algorithm used in a study of subjects with diabetes. de Bresser et al[21] used the same algorithm in a separate study of subjects with diabetes and white matter lesion load. Tiehuis et al[22] also used the same algorithm in a study of cognitive function, vascular disease and diabetes, and, in another study, showed that it performs favorably compared with subjective rating scales with reference to cognitive assessments.[23] For all of these studies, performance of the described algorithm was previously evaluated by using a leave-on-out cross-validation procedure, achieving a similarity index of 0.8.[24] This validation, however, was performed only on 10 elderly subjects with a history of vascular disease and not specifically on patients with diabetes. A recently described method with the use of support vector machines was evaluated on a subset of 45 subjects from a larger study on the treatment of diabetes.[25] A single rater was used as the reference standard, with 10 subjects used for training and the remaining 35 for testing. This method demonstrated a high sensitivity (>0.9) and a specificity of ~0.85. Interestingly, this study had a second manual rater but only used the first rater as the reference

standard, possibly because of their reported large inter-rater variability. In comparison to these studies, our study provides a more direct evaluation of lesion segmentation in diabetes, including a large sample size, and a more rigorous approach to ground-truth determination.

WM Disease and Diabetes

The LST performance was very similar between DM-affected and unaffected individuals, achieving similar maximum correlations and optimum k-thresholds. There was no significant difference in degree of WM lesion load between affected and unaffected groups, by use of any of the metrics (WM scores, LST, or GT). Whereas there is a convincing relationship between DM and lacunar infarcts and brain atrophy, the association with WM hyperintensities on conventional MR imaging is less clear.[26] There has been recent evidence suggesting that DM is an independent risk factor for deep WM lesions in the elderly.[27] Additionally, voxelwise analyses of diffusion tensor imaging and fractional anisotropy have demonstrated early changes in white matter microstructural properties that were associated with diabetes duration.[28] Our sample may not have been large enough to detect a between-group difference. Alternatively, the effects of aging may be a greater contributor to the presence of visible WM disease than DM. It should be noted that our DM population had very few subjects with WM scores >5, and none were >7. We drew a random sample from the Diabetes Heart Study, with WM disease burdens probably reflecting the distribution of WM disease in the general DM population. Conclusions about diabetes and WM disease should not be made on the basis of this study because we did not control for a variety of confounding effects, including disease duration, co-morbidities, and medications.

Limitations

No subject had WM scores >7 or TLV >35 mL. This limits evaluation at the extreme end of disease burden. Also, our population was skewed toward the lower end of disease burdens, with most having TLV <1 mL. Although this is a limitation in terms of disease distribution, it is also probably reflective of the true disease burden and incidence of WM disease in the general population. One critical limitation is generic to all studies of WM disease burden, which is that there is no accepted reference standard. Manual delineation by experienced readers comes closest to what most would accept as a reference standard. The laborious nature of manual WM segmentation makes it generally impractical to have multiple raters. Semi-automated methodologies to facilitate the procedure can introduce bias through predetermination of potential lesion borders. Even with multiple readers, some readers are typically more accurate (or meticulous). These limitations may be mitigated by machine learning approaches that can weight more accurate readers more heavily in the determination of GT. We attempted to address these potential biases in GT segmentation with the use of a multi-tiered manual segmentation approach with the final consensus manual segmentation performed by experienced neuroradiologists. This is a time-intensive and laborious approach but allows confidence in the quality of the GT determination.

CONCLUSIONS

We validate the use of the LST for determination of TLV in a diabetic population and demonstrate that it performs as well compared with GT as a widely used subjective WM lesion rating scale. Additionally, we identify an optimal k-threshold of 0.25, with robust performance between 0.2–0.4 in this population. The LST is a readily available substitute for subjective WM lesion scoring in studies of diabetes and other populations prone to leukoaraiosis. Studies that use subjective WM lesion scores should be cognizant of violations in assumptions of linearity for standard statistical models with the use of this scale.

REFERENCES

1. Brown WR, Moody DM, Thore CR, et al. **Apoptosis in leukoaraiosis.** *AJNR Am J Neuroradiol* 2000;21:79–82
2. Murray AD, Staff RT, Shenkin SD, et al. **Brain white matter hyperintensities: relative importance of vascular risk factors in nondemented elderly people.** *Radiology* 2005;237:251–57
3. Debette S, Markus HS. **The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis.** *BMJ* 2010;341:c3666
4. Prins ND, van Dijk EJ, den Heijer T, et al. **Cerebral white matter lesions and the risk of dementia.** *Arch Neurol* 2004;61:1531–34
5. Kuller LH, Longstreth WT Jr, Arnold AM, et al. **White matter hyperintensity on cranial magnetic resonance imaging: a predictor of stroke.** *Stroke* 2004;35:1821–25
6. Debette S, Beiser A, DeCarli C, et al. **Association of MRI markers of vascular brain injury with incident stroke, mild cognitive impairment, dementia, and mortality: the Framingham Offspring Study.** *Stroke* 2010;41:600–06
7. Manolio TA, Kronmal RA, Burke GL, et al. **Magnetic resonance abnormalities and cardiovascular disease in older adults: the Cardiovascular Health Study.** *Stroke* 1994;25:318–27
8. Yue NC, Arnold AM, Longstreth WT Jr, et al. **Sulcal, ventricular, and white matter changes at MR imaging in the aging brain: data from the Cardiovascular Health Study.** *Radiology* 1997;202:33–39
9. Schmidt P, Gaser C, Arsic M, et al. **An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis.** *Neuroimage* 2012;59:3774–83
10. Bowden DW, Cox AJ, Freedman BI, et al. **Review of the Diabetes Heart Study (DHS) family of studies: a comprehensively examined sample for genetic and epidemiological studies of type 2 diabetes and its complications.** *Rev Diabet Stud* 2010;7:188–201
11. Divers J, Hugenschmidt C, Sink KM, et al. **Cerebral white matter hyperintensity in African Americans and European Americans with type 2 diabetes.** *J Stroke Cerebrovasc Dis* 2012 May 16 [Epub ahead of print]
12. Ashburner J, Friston KJ. **Voxel-based morphometry: the methods.** *NeuroImage* 2000;11:805–21
13. Rorden C, Karnath HO, Bonilha L. **Improving lesion-symptom mapping.** *J Cogn Neurosci* 2007;19:1081–88
14. Fisher RA. **Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population.** *Biometrika* 1915;10:507–21
15. Steiger JH. **Tests for comparing elements of a correlation matrix.** *Psychol Bull* 1980;87:245–51
16. Filippi M, Rocca MA, Barkhof F, et al. **Association between pathological and MRI findings in multiple sclerosis.** *Lancet Neurol* 2012;11:349–60

17. Brown WR, Moody DM, Challa VR, et al. **Apoptosis in leukoaraiosis lesions.** *J Neurol Sci* 2002;203–204:169–71

18. Grueter BE, Schulz UG. **Age-related cerebral white matter disease (leukoaraiosis): a review.** *Postgrad Med J* 2012;88:79–87

19. Knottnerus IL, Ten Cate H, Lodder J, et al. **Endothelial dysfunction in lacunar stroke: a systematic review.** *Cerebrovasc Dis* 2009;27:519–26

20. Jongen C, van der Grond J, Kappelle LJ, et al. **Automated measurement of brain and white matter lesion volume in type 2 diabetes mellitus.** *Diabetologia* 2007;50:1509–16

21. de Bresser J, Tiehuis AM, van den Berg E, et al. **Progression of cerebral atrophy and white matter hyperintensities in patients with type 2 diabetes.** *Diabetes Care* 2010;33:1309–14

22. Tiehuis AM, Mali WP, van Raamt AF, et al. **Cognitive dysfunction and its clinical and radiological determinants in patients with symptomatic arterial disease and diabetes.** *J Neurol Sci* 2009;283:170–74

23. Tiehuis AM, Vincken KL, Mali WP, et al. **Automated and visual scoring methods of cerebral white matter hyperintensities: relation with age and cognitive function.** *Cerebrovasc Dis* 2008;25:59–66

24. Anbeek P, Vincken KL, van Bochove GS, et al. **Probabilistic segmentation of brain tissue in MR imaging.** *Neuroimage* 2005;27:795–804

25. Lao Z, Shen D, Liu D, et al. **Computer-assisted segmentation of white matter lesions in 3D MR images using support vector machine.** *Acad Radiol* 2008;15:300–13

26. van Harten B, de Leeuw FE, Weinstein HC, et al. **Brain imaging in patients with diabetes: a systematic review.** *Diabetes Care* 2006;29:2539–48

27. van Harten B, Oosterman JM, Potter van Loon BJ, et al. **Brain lesions on MRI in elderly patients with type 2 diabetes mellitus.** *Eur Neurol* 2007;57:70–74

28. Hsu JL, Chen YL, Leu JG, et al. **Microstructural white matter abnormalities in type 2 diabetes mellitus: a diffusion tensor imaging study.** *Neuroimage* 2012;59:1098–105