The Siena/FSL whole brain atrophy measurement algorithm may require substantially larger group sizes at 3T than 1.5T for Alzheimer's disease

K. S. Cover[1], R. A. van Schijndel[2], B. W. van Dijk[3], A. Redolfi[4], D. L. Knol[5], G. B. Frisoni[4], F. Barkhof[2], and H. Vrenken[2,6]

[1]Physics and Medical Technology, VU University Medical Center, Amsterdam, North Holland, Netherlands, [2]Department of Radiology, VU University Medical Center, Amsterdam, Netherlands, [3]Department of Physics and Medical Technology, VU University Medical Center, Amsterdam, Netherlands, [4]Laboratory of Epidemiology & Neuroimaging, IRCCS San Giovanni di Dio Fatebenefratelli, Brescia, Italy, [5]Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, Netherlands, [6]MS Center Amsterdam and Alzheimer Center, VU University Medical Center, Amsterdam, Netherlands

Purpose

To compare the back-to-back (BTB) reproducibility of the Siena/FSL whole brain atrophy measurement algorithm for MPRAGEs acquired at 3T and 1.5T on the same subject.

Background

MRI based measurement of atrophy, such as whole brain and hippocampus, are being used as biomarkers in diseases such as multiple sclerosis (MS) and Alzheimer's disease. There is substantial evidence to demonstrate that whole brain atrophy measurements perform significantly different at 3T than 1.5T [1]. BTB reproducibility studies have been conducted at 1.5T previously [2-3]. However, no BTB studies have been conducted at 3T. The Alzheimer's Disease Neuroimaging Initiative (ADNI) is scanning hundreds of subjects at both 3T and 1.5T providing a valuable benchmark against which to access the reproducibility of atrophy measurement algorithms such as Siena/FSL.

Method

BTB MPRAGE's at both 3T and 1.5T were downloaded from the ADNI website for the month 0 and month 12 scans. A total of 118 subjects, drawn from all 3 of ADNI's diagnostic groups and scanned before February 2010, had all 8 of the required MPRAGEs available for analysis. Siena/FSL (version FSL version 4.1.4) was used to calculate the percentage brain volume change (PBVC) between month 0 and month 12 for each of the BTB MPRAGEs. The BTB difference for each subject was defined to be the PBVC of the first of the pair of MPRAGEs acquired minus that of the second. In the ADNI study, the second BTB MPRAGE starts within a few seconds of the completion of the first giving the subject little opportunity to move.

To assess the reproducibility over various groups, several spread statistics of the BTB difference were calculated. These included the commonly used 50 percentile of the absolute value of the BTB difference [2], the less commonly used 90 percentile of the absolute value of the BTB difference and the standard deviation of the BTB difference (Table 1). A scatter plot of the 3T and 1.5T BTB differences was generated.

As an additional measurement of the reproducibility, a power calculation to determine the group size required to detect a specific treatment effect was completed. A bootstrap simulation based on the Wilcoxon-Mann-Whitney test was implemented to estimate the group size thus avoiding any potential errors introduced by assuming a Gaussian distribution.

Results and Discussion

Visual inspection of the scatter plot in Figure 1 shows similar BTB differences for most subjects at 3T and 1.5T. Although the numbers are small, there seems to be no visually obvious differences between 3T and 1.5T or among the 3 diagnostic groups.

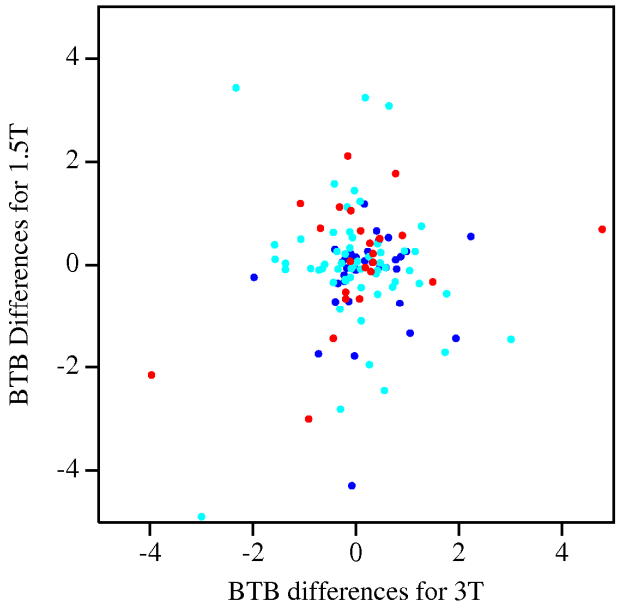

Figure 1 Scatter graph of 3T versus 1.5T reproducibility for subjects in the ADNI data set. HC (blue), MCI (light blue), AD (red)

Table 1 Spread statistics for BTB reproducibility for both 3T and 1.5T

BTB Difference Reproducibility	Combined		HC		MCI		AD	
	3T	1.5T	3T	1.5T	3T	1.5T	3T	1.5T
N	118		37		58		23	
50 percentile	0.39	0.35	0.40	0.26	0.42	0.33	0.33	0.66
90 percentile	1.61	1.90	1.94	1.44	1.58	2.45	3.97	2.15
Standard Dev	1.33	1.14	1.69	0.93	0.96	1.27	1.44	1.17

The statistics presented in Table 1 lead to a similar conclusion as the scatter plot. The 50 percentile, 90 percentile and standard deviation for the combined diagnostic groups also yield similar results for the 3T and 1.5T MPRAGEs. While the reproducibility statistics are also broken down by diagnostic groups, the small number of subjects in some statistics makes the statistics for the diagnostic groups less reliable than the combined group. This is especially true for AD.

The group sizes (Table 2) present a different picture than the scatter plot or BTB statistics. To detect a specified percentage point reduction in the annual progression of the atrophy in either MCI or AD patients, the 3T MPRAGEs consistently requires roughly 50% larger groups. In table 2 a percentage point reduction of about 0.6% would correspond to a slowing of about 50% in the progression of AD. Also, roughly a 0.25% percentage point reduction would correspond to a 50% slowing in the progression of MCI. Although, in both cases the amount of slowing is cohort dependent.

Table 2 Group sizes for treatment effects

Percentage Point Reduction	Group Size			
	MCI		AD	
	3T	1.5T	3T	1.5T
1.0	49	32	50	29
0.8	77	48	75	45
0.6	131	81	129	77
0.4	291	180	283	172

The non Gaussian nature of the BTB difference distribution, as characterized by much broader shoulders than a Gaussian distribution [3], demands more care in the interpretation of the results. This non Gaussian distribution may be why Smith et al. [2], instead of using the much more common standard deviation to quantify the spread of the BTB difference, choose to use the 50 percentile of the absolute value of the BTB difference. While for a Gaussian distribution the spread statistics in Table 1 can be a good indication of the difference between the two groups, this does not hold in general. The more robust and reliable prediction of the performance of Siena in a clinical trial is the bootstrap simulation based on a non parametric statistical test – as was used to calculate the group size in this abstract.

While it would be desirable for this result on the required group sizes at 3T and 1.5T to be confirmed both with larger numbers and other data sets, it should be helpful to keep the result in mind during design of clinical trials. Other atrophy measurement algorithms also need to be similarly assessed.

Referemces [1] Keihaninejad S et al. NeuroImage. 2010:50;1427–1437. [2] Smith SM et al. NeuroImage 2007;36:1200-1206. [3] Cover et al. 2010. Psychiatry Research: Neuroimaging Under revision.