Neuroscience Campus Amsterdam

A standard benchmark for assessing the reproducibility of brain atrophy measures in Alzheimer's using the ADNI1 data set

Keith S Cover[a], Ronald A van Schijndel[a], Adriaan Versteeg[a], Alberto Redolfi[b], Jérôme Revillard[c], Baptiste Grenier[c], David Manset[c], Hugo Vrenken[a], Bob W van Dijk[a], Giovanni B Frisoni[b], Frederik Barkhof[a], neuGRID Consortium

[a]VU University Medical Center, Amsterdam [b]IRCCS San Giovanni di Dio Fatebenefratelli, Brescia, Italy [c]MAAT, Archamps, France (Contact: Keith@kscover.ca)

Purpose

To compare the reproducibility of the hippocampal atrophy rates over one year generated by the fully automated FreeSurer/ReconAll and FSL/FIRST software packages using a back-to-back (BTB) reproducibility test based on the ADNI1 data set [1] .

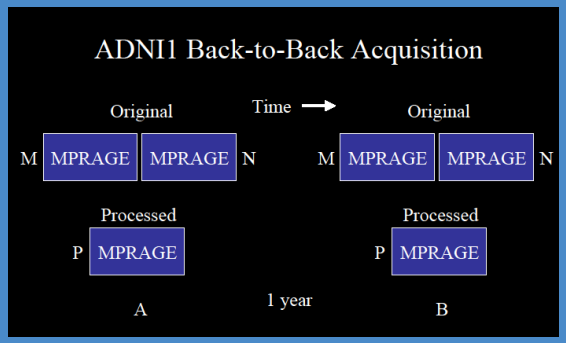

Figure 1. Acquisition protocol for the MPRAGE MRI scans in the ADNI1 study. For each patient visit ADNI selected one of the pair of original scans acquired for processing and official release.

Methods

FSL/FIRST [1] (fsl.fmrib.ox.ac.uk/fsl/fslwiki) and FreeSurfer/ReconAll (surfer.nmr.mgh.harvard.edu) are widely used as fully automatic measures of hippocampal atrophy from a pair of 3D T1 weighted MRI scans. Back-to-back (BTB) ADNI1 MPRAGE acquired at baseline (A) and 1 year (B) were used to calculate the BTB difference between the first acquired MPRAGE (M) at each patient visit from the second acquired (N). After taking the absolute value of the BTB differences, the 50 percentile was found and used as a statistic of reproducibility of the BTB difference.

The full ADNI1 BTB benchmark, which was designed for this study, consisted of 562 subjects each with 6 MPRAGEs. The 6 MPRAGEs each consisted of 3 MPRAGEs of two patient visits at baseline and 1 year. Each patient visit consisted of the two BTB "original" MPRAGEs and a third "processed" MPRAGE (P) - which is one of the two original MPRAGEs with additional processing by ADNI.

As the computational time for ReconAll was about 80 times longer than for FIRST, a representative subset of 75 subjects was used for each of the 4 versions of FreeSurfer benchmarked. FIRST was calculated for both the 75 and the full 562.

When, occasionally, a subject's calculation yielded no output, the N value was correspondingly decreased. All FreeSurfer versions were run on the same cluster.

ADNI1 Back-to-Back MPRAGEs

While rarely mentioned in the literature, as part of the first Alzheimer's Disease Neuroimaging Initiative (ADNI1) study, the 3D T1-weighted MRI scans (also know as MPRAGE scans) were acquired in duplicate during each patient visit - with the acquisition of the second MPRAGE starting within seconds of completion of the first. As ADNI has over 800 subjects - with an average of 6 visits each – spread over several years, roughly 9,000 back-to-back (BTB) MPRAGE were available to probe the performance of brain atrophy measures [1].

Reproducibility of Hippocampal Atrophy Rates

Algorithm	Cross Sectional		Longitudinal		N
	Left	Right	Left	Right	
FreeSurfer 4.5.0	2.41%	2.06%	2.01%	1.73%	75
FreeSurfer 5.0.0	2.80%	3.20%	**13.98%**	**13.89%**	75-72
FreeSurfer 5.1.0	2.38%	2.77%	2.31%	2.10%	75
FreeSurfer 5.2.0	3.29%	2.93%	2.36%	1.71%	75
FSL 5.0.4	2.19%	3.05%	N/A	N/A	74
FSL 5.0.4	2.47%	2.87%	N/A	N/A	557

Table 1. The reproducibility of hippocampal atrophy rates for FreeSurfer/ReconAll and FSL/FIRST using back-to-back (BTB) ADNI1 MPRAGEs MRI scans . Units are percentage difference in the atrophy rate over 1 year. Smaller is better.

Results

Table 1 presents the BTB reproducibility for FLS 5.0.4 and several versions of FreeSurfer.

Comparison of FreeSurfer 5.2.0 and FSL 5.0.4 shows similar reproducibilities with FSL being slightly better for the left hippocampus and FreeSurfer 5.2.0 being slightly better on the right hippocampus.

The range of reproducibilities of FreeSurfer in cross sectional mode over the various versions gives a sense of the variation in the reproducibility statistic. FIRST and the 4 versions of ReconAll do not exhibit significant difference reproducibility of the left and right hippocampus.

As the 75 subjects in the subset is a representative sample of the full 562 subjects in the benchmark, repeating the calculations for the full 562 is unlikely to yield substantially different results.

While all the results are calculated here for the fully automated segmentation, it would be interesting to see if manual review of the segmentation changed the reproducibilities significantly.

Conclusions

- The FSL/FIRST and FreeSurfer/ReconAll reproducibilities of hippocampal atrophy rates are similar.
- The poor reproducibility of the well documented bug in longitudinal mode of FreeSurfer 5.0.0 is clearly evident.
- The standard BTB benchmark presented - based on ADNI1 data - for measuring the reproducibility of brain atrophy measures is a valuable way to compare performance of algorithms

References: [1] Cover KS, et al. Psychiatry Research: Neuroimaging. 2011;193:182.

Study funding was provided by neuGRID4you (N4U), an European Community FP7 project (grant agreement 283562), and the VU University Medical Center, Amsterdam, The Netherlands

SEVENTH FRAMEWORK PROGRAMME

neuGRID 4U

VU medisch centrum

vrije Universiteit *amsterdam*