

Validation of the EADC-ADNI Harmonized Protocol for Manual Hippocampal Segmentation: Preliminary Results

Marina Boccardi, Martina Bocchetta, Frederik Barkhof, Corinna Bauer, Melanie Blair, Claire Boutet, Emma Burton, Enrica Cavedo, Adam Christensen, Kristian Steen Frederiksen, Michel Grothe, Sarah Hollander, Mariangela Lanfredi, Yawu Liu, Oliver Martinez, Masami Nishikawa, Marileen Portegies, Gregory Preboske, Travis Stoub, Tim Swihart, Mat Tinley, Chad Ward, Clarissa Ferrari, Patrizio Pasqualetti, Simon Duchesne, Clifford R. Jack Jr., Giovanni Frisoni

Background

A Harmonized Protocol (HP) for manual hippocampal segmentation was defined by a Delphi panel (Boccardi et al., *Neurology* 2012;78(S1):S04003), described in a written document, implemented for standard learning by new tracers, and submitted to validation.

Methods

Twenty ADNI magnetic resonance images (MRIs) were selected for being balanced by hippocampal atrophy severity at Scheltens' scale (0–4) and magnet field strength (1.5T, 3T). Twenty-one tracers with ICC>0.80 in hippocampal segmentation based on local protocols were recruited from independent centres to segment the right and left hippocampi based on local protocols, to learn segmentation based on the HP and to re-segment the same images based on the HP. The same version of MultiTracer was used for manual segmentation in each phase. The 10 tracers who completed the HP training and the re-segmentation had high Jaccard indices versus benchmark segmentations (range: 0.78–0.85, Mdn: 0.80). Volume ICCs were computed with both the Consistency and Absolute methods, and compared with t-tests.

Results

Twenty-one tracers completed segmentations based on local protocols. Inter-rater ICCs were as follows. Consistency, left: 0.826 (C.I.: 0.728–0.912), right: 0.807 (0.702–0.901), Absolute, left: 0.459 (0.274–0.672), right: 0.434 (0.254–0.649). Of these tracers, 10 completed the re-segmentation based on the HP. ICCs computed based on local protocols were similar in this sub-sample (see Table and Figure). ICCs based on the HP were: Consistency, left: 0.966 (0.939–0.984), right: 0.971 (0.949–0.987); Absolute, left: 0.875 (0.719–0.948), right: 0.876 (0.711–0.949). The difference of these higher ICCs from those of local protocols was highly significant at t-test (up to $p<0.0005$, see Table).

Conclusions

Segmentations based on the HP achieved higher inter-rater reliability values, with lower confidence intervals, than those performed based on local protocols. Inter-rater reliability across tracers from different laboratories and using the most restrictive “Absolute” method were very high. These results are preliminary, the completion of this Harmonization phase is planned for June 2013.