The use hippocampal volumetry as a biomarker of Alzheimer's Disease (AD) requires standard operating procedures. A Delphi panel of experts converged on a Harmonized Protocol (HP) for manual segmentation from magnetic resonance images (MRIs).

To produce benchmark images of hippocampal segmentation reflecting the HP, to be released publicly as the gold standard for human tracers and algorithms.

Two rounds of corrections were asked to tracers. One round of corrections was applied to the HP. Reliability values for the corrected segmentations were: lowest absolute 5-level intra-rater, ICC 0.943 (95% CI 0.335-0.989); lowest inter-rater: ICC 0.943 (0.791-0.986) (Table 2). The mean 5-level inter-rater values were ICC 0.96 (absolute) and ICC 0.98 (consistency). Overlapping reliability among the 5 tracers was 0.73 for 1.5T and 0.75 for 3T images (Figures 1 and 2).

The HP showed to produce very reliable manual segmentations. The obtained hippocampal segmentations appear as an appropriate benchmark set for certification of tracers who will carry out the validation of the HP.