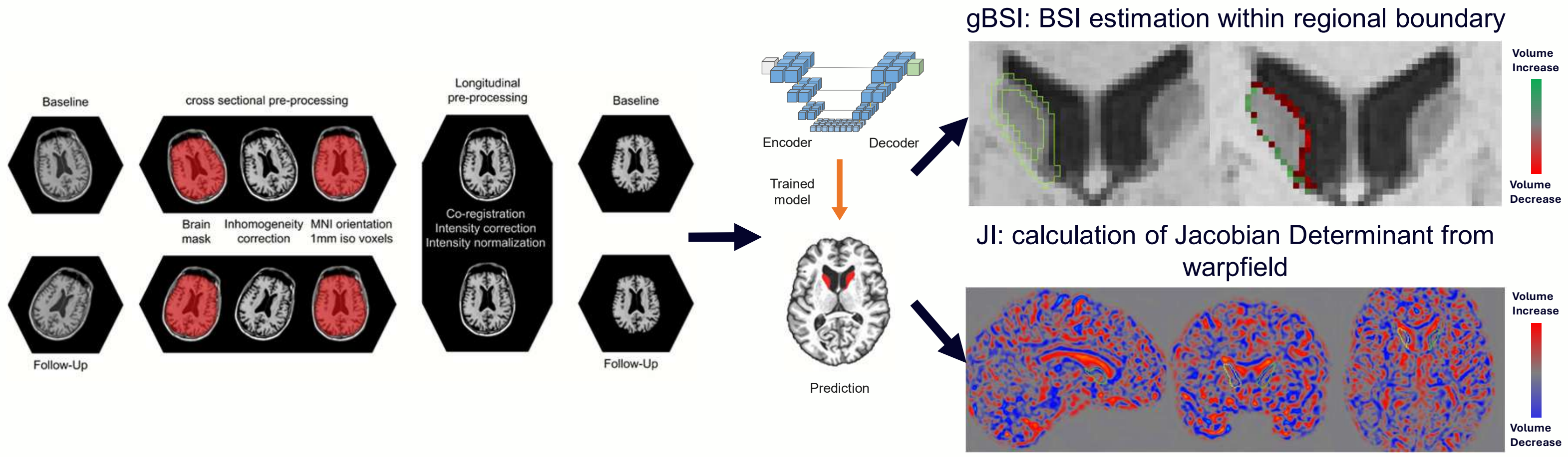
Comparison of automated methods for volumetric analyses in Huntington's disease against the current best practice, the Boundary Shift Integral with manual segmentation

Michelle Lax[1], Marina Papoutsi[1], Robin Wolz[1,2]
1 IXICO, London UK |2 Imperial College London UK

The current best practice for measuring atrophy in Huntington's disease (HD), validated in Track-HD [1], requires manual segmentation of caudate and whole brain regions combined with the Boundary Shift Integral (BSI) for measurement of longitudinal change by evaluating the change in the intensity gradient within the regional boundaries [2]. This approach, however, is labour intensive and requires significant analyst expertise.

We have developed a fully automated framework that uses deep learning for regional segmentation (IXIQ.Ai; [3]) and can be combined with generalised BSI (gBSI; [4]) or Jacobian Integration (JI; [5]) methods for longitudinal measurement. IXIQ.Ai-gBSI uses the IXIQ.Ai probabilistic mask for the definition of the regional boundary, whereas IXIQ.Ai-JI is a deep-learning voxel-wise non-linear registration method estimating volume change from the determinant of the warpfield. Here, we compared the IXIQ.Ai-gBSI and -JI methods against the current best practice (manual-BSI).

Image Pre-processing | *Regional Segmentation* | *Boundary Shift Integral*



Methods

We retrospectively analysed 60 T1-weighted MRI images from Track-HD: 30 HD gene-carriers prior to clinical diagnosis (preHD; CAG>=39 and diagnostic confidence level (DCL)<4), and 30 with a clinical motor diagnosis (HD; CAG>=39 and DCL=4). Details are provided in the Table below. All data were acquired using 3T scanners. Baseline and 1-year follow-up images were analysed using manual-BSI, IXIQ.Ai-gBSI and IXIQ.Ai-JI.

Demographics	preHD	HD
N	30	30
Years from Baseline at Follow-Up - Mean (SD)	0.98 (0.11)	0.97 (0.09)
Age - Mean (SD)	41.30 (7.96)	49.63 (9.42)
Sex (%M)	56.67%	46.67%
CAG - Median (SD)	43 (2.19)	43 (2.33)

We used separate linear mixed models for each method with random intercept and factors for time (baseline and 1 year), group (preHD and HD), age, sex, and their interaction with time.

The outcome measures were caudate and whole brain volumes normalised by the affine scaling factor to correct for head size. P-values were Bonferroni corrected for multiple comparisons (six tests). Cohen's d was used to quantify effect size (defined as mean estimated difference divided by residual SD), where larger values mean more sensitive to change.

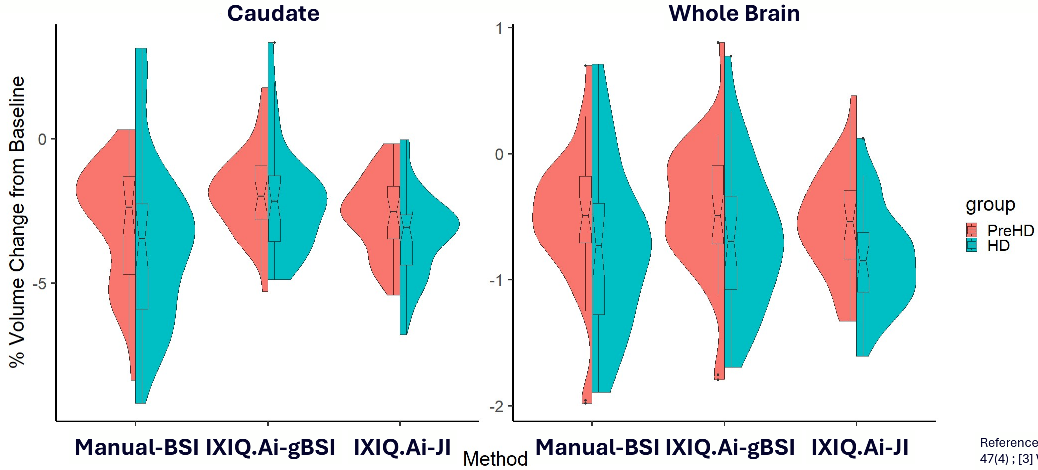
Results

The main effects of group and time were significant in all cases (all corrected p-values<0.001), but there was no significant time by group interactions for any of the comparisons (all corrected p-values>0.05). As such there was no difference between groups in terms of rate of change.

The fully automated IXIQ.Ai-JI method had larger effect sizes for both the caudate and whole brain compared to both the manual-BSI and the IXIQ.Ai-gBSI methods. In addition, the fully automated IXIQ.Ai-gBSI performed very similar to the manual-BSI method.

Method	Main Effect of Time (Change from Baseline)	Whole Brain	Caudate
Manual-BSI	Contrast Mean (95%CI) (mm^3)	8595.7 (6385.6-10805.7)	180.8 (145.1-216.4)
	P-value	<0.001	<0.001
	Cohen's d (95%CI)	1.47 (1.04-1.90)	1.92 (1.46-2.37)
IXIQ.Ai-gBSI	Contrast Mean (95%CI) (mm^3)	7528.0 (5599.7-9696.0)	119.7 (97.0-148.5)
	P-value	<0.001	<0.001
	Cohen's d (95%CI)	1.40 (0.98-1.82)	1.73 (1.29-2.18)
IXIQ.Ai-JI	Contrast Mean (95%CI) (mm^3)	9314.6 (7945.1-10684.0)	176.2 (154.6-197.7)
	P-value	<0.001	<0.001
	Cohen's d (95%CI)	2.57 (2.06-3.08)	3.09 (2.53-3.65)

Figures



Conclusion

The violin plots show % change from baseline for the three methods. IXIQ.Ai-JI has the smallest variability for both ROIs, which drives the differences in effect size and sensitivity to change.

In conclusion, compared to current best practice (manual-BSI), our fully automated IXIQ.Ai framework performed better when combined with JI and had comparable performance when combined with gBSI. However, IXIQ.Ai solutions are faster, less labour intensive, and scalable making them better suited for clinical development.

References: [1] Tabrizi et al., 2013 Lancet Neurology, 12 (7); [2] Hobbs et al., 2009, Neuroimage, vol. 47(4) ; [3] Weatheritt et al., 2020, Neurotherapeutics, 17(S1); [4] Prados et al., Neurobiology of Aging, 2015, 36; [5] Reinwalk et al., 2021, Alzheimer's & Dementia, e050068