COMPARING FREESURFER WITH MANUAL SEGMENTATION IN THE BASAL GANGLIA AND THALAMUS OF 7 YEAR OLD CHILDREN

Wai Yen Loh[1,2], Zohra M Ahmadzai[2,3], Lillian Gabra Fam[2,3], Alan Connelly[1], Alicia J Spittle[2,4], Katherine J Lee[2,3], Terrie E Inder[5], Jeanie LY Cheong[2,4], Lex W Doyle[2,4], Peter J Anderson[2,3], and Deanne K Thompson[1,2]

[1]*The Florey Institute of Neuroscience and Mental Health, Melbourne, VIC, Australia,* [2]*Murdoch Childrens Research Institute, Melbourne, VIC, Australia,* [3]*Department of Paediatrics, University of Melbourne, Melbourne, VIC, Australia,* [4]*Royal Women's Hospital, Melbourne, VIC, Australia,* [5]*Brigham and Women's Hospital, Boston, MA, United States*

Background & Aim: Although considered the 'gold standard' for obtaining brain structure volumes, manual segmentation of subcortical structures is labour-intensive and time-consuming. As a result, automated and semi-automated segmentation methods have been developed to quantify subcortical volumes. FreeSurfer is a commonly used brain segmentation software that can be operated as fully-automated or semi-automated.[1] However, current literature on the validity of FreeSurfer compared with manual segmentation is limited in paediatrics; especially for the basal ganglia and thalamus which are important structures known to be affected by preterm birth.[2] The aim of this study was to validate the reliability of FreeSurfer's semi-automated segmentation results of the basal ganglia and thalamus against manual segmentation in a paediatric cohort using absolute agreement and consistency measures. The two methods are in absolute agreement if the volumes from the two approaches match exactly, whereas the methods are consistent if the volumes from the two approaches differ by a constant value for all measurements.

Methods: At 7 years of age, 20 term controls and 20 very preterm children (<30 weeks' gestation) underwent magnetic resonance imaging. 3D T1-weighted images were acquired with the following sequence parameters: TR= 1900ms, TE= 2.27ms, Flip angle= 9°, FOV= 210 x 210mm, Matrix= 256 x 256, 0.8mm^3 isotropic voxels. FreeSurfer v4.4 was used to automatically segment the whole brain, including the structures of interest: left and right caudate, putamen, nucleus accumbens, globus pallidus and thalamus (Figure 1a). FreeSurfer output was manually edited if there were obvious errors and this was performed by two raters. Inter-rater reliability was assessed using intra-class correlation coefficients (ICC: from a two-way mixed models) for absolute agreement and ICCs were >0.9 for all structures (ranged between 0.936 and 0.995). For the manual method, the same 40 images were first aligned to the anterior commissure and posterior commissure axis using 3D Slicer v4.1.1. Each structure was then manually segmented by a single operator on the coronal view and edited in the other orthogonal views with ITK-SNAP v2.2.0 using previously established guidelines (Figure 1b). [3-7] To reduce the possibility of handedness bias, 20 scans were flipped on the x-axis (coronal view). To assess intra-rater reliability of the single operator, manual segmentation was repeated in 10 images. ICCs for absolute agreement were >0.8 for all structures (ranged between 0.847 and 0.988). The semi-automated and manual methods were compared using the ICC (absolute agreement and consistency) and Bland-Altman plots.

Results: The ICCs for absolute agreement between FreeSurfer and manual segmentation ranged from 0.265 to 0.829. In contrast, the ICCs for consistency between the two methods were >0.7 for all structures (Table 1), except the right caudate (0.693) and nucleus accumbens (right: 0.614, left: 0.307). Bland-Altman plots indicated that when compared with manual segmentation, FreeSurfer volumes were in general smaller for the caudate, but larger for the putamen, thalamus, nucleus accumbens and left globus pallidus. There was no clear bias observed in the right globus pallidus.

Discussion & Conclusions:
Although the basal ganglia and thalamus volumes obtained by FreeSurfer and manual segmentation had poor to average ICCs for absolute agreement in all structures except the globus pallidus and left caudate, the ICCs for consistency were relatively high in all structures except the nucleus accumbens. The poor consistency for ICCs in the nucleus accumbens is most likely due to the manual tracing protocol being solely based on anatomical landmarks whenever there was overlap between the nucleus accumbens and other structures as opposed to the FreeSurfer method which uses both spatial and voxel intensity information in its segmentation. Results from the Bland-Altman plots suggest that there is a systematic bias between FreeSurfer and the 'gold standard' manual segmentation. In particular, FreeSurfer typically underestimated the volume of the caudate but overestimated the volume of the putamen, nucleus accumbens, thalamus and left globus pallidus. Although these results indicate that FreeSurfer may not provide similar basal ganglia and thalamus volumes to those of manual segmentation, it may still be useful in comparing subcortical volume differences between healthy and diseased paediatric groups. Nonetheless, if studies require accurate basal ganglia and thalamus volumes, manual segmentations are recommended. The results of this study will be useful for informing other studies wishing to investigate volumes of the basal ganglia and thalamus in paediatric populations.

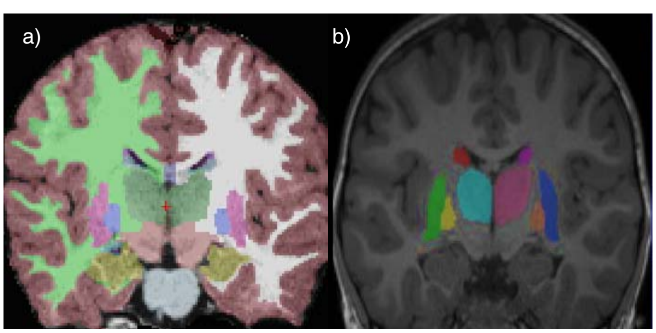

Figure 1. **(a)** FreeSurfer (left) and **(b)** manual (right) segmentations of the basal ganglia and thalamus.

Table 1. Comparing semi-automated FreeSurfer (FS) with manual segmentation

Brain Structure	Intraclass Correlation Coefficient		Mean Volume Difference (mm^3) [FS-Manual]	
	Absolute Agreement	Consistency	Mean Difference	Std. Dev.
Right Caudate	0.631	0.693	-298.6	507.15
Left Caudate	0.721	0.740	-166.6	473.6
Right Putamen	0.554	0.902	745.1	292.7
Left Putamen	0.409	0.898	1040.5	303.8
Right Nucleus Accumbens	0.292	0.614	194.7	114.5
Left Nucleus Accumbens	0.265	0.307	72.0	142.8
Right Globus Pallidus	0.829	0.825	-3.2	147.6
Left Globus Pallidus	0.716	0.791	117.2	160.6
Right Thalamus	0.483	0.813	838.6	438.1
Left Thalamus	0.573	0.865	768.2	393.8

References:
1 Fischl, B *et al.* Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain. *Neuron* **33**, 341-355 (2002).
2 Srinivasan, L *et al.* Quantification of Deep Gray Matter in Preterm Infants at Term-Equivalent Age Using Manual Volumetry of 3-Tesla Magnetic Resonance Images. *Pediatrics* **119**, 759-765 (2007).
3 Westmoreland, P *et al. Caudate Tracing Guidelines*, www.icts.uiowa.edu/confluence/download/attachments/72385476/caudate.pdf.
4 Westmoreland, P *et al. Putamen Tracing Guidelines*, www.icts.uiowa.edu/confluence/download/attachments/72385476/putamen.pdf.
5 Ootemann, W *et al. Thalamus Tracing Guidelines*, www.icts.uiowa.edu/confluence/download/attachments/72385476/thalamus.pdf.
6 Langen, M *et al.* Caudate nucleus is enlarged in high-functioning medication-naive subjects with autism. *Biological Psychiatry* **62**, 262-266 (2007).
7 Brambilla, P *et al.* Anatomical MRI study of basal ganglia in bipolar disorder patients. *Psychiatry Research: Neuroimaging* **106**, 65-80 (2001).