

Neuroinform (2012) 10:341–350
DOI 10.1007/s12021-012-9147-0

ORIGINAL ARTICLE

Volume Estimation of the Thalamus Using FreeSurfer and Stereology: Consistency between Methods

Simon S. Keller · Jan S. Gerdes ·
Siawoosh Mohammadi · Christoph Kellinghaus ·
Harald Kugel · Katja Deppe · E. Bernd Ringelstein ·
Stefan Evers · Wolfram Schwindt · Michael Deppe

Published online: 6 April 2012

© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract Freely available automated MR image analysis techniques are being increasingly used to investigate neuro-anatomical abnormalities in patients with neurological disorders. It is important to assess the specificity and validity of automated measurements of structure volumes with respect to reliable manual methods that rely on human anatomical expertise. The thalamus is widely investigated in many

neurological and neuropsychiatric disorders using MRI, but thalamic volumes are notoriously difficult to quantify given the poor between-tissue contrast at the thalamic gray-white matter interface. In the present study we investigated the reliability of automatically determined thalamic volume measurements obtained using FreeSurfer software with respect to a manual stereological technique on 3D T1-weighted MR images obtained from a 3 T MR system. Further to demonstrating impressive consistency between stereological and FreeSurfer volume estimates of the thalamus in healthy subjects and neurological patients, we demonstrate that the extent of agreeability between stereology and FreeSurfer is equal to the agreeability between two human anatomists estimating thalamic volume using stereological methods. Using patients with juvenile myoclonic epilepsy as a model for thalamic atrophy, we also show that both automated and manual methods provide very similar ratios of thalamic volume loss in patients. This work promotes the use of FreeSurfer for reliable estimation of global volume in healthy and diseased thalami.

S. S. Keller · J. S. Gerdes · E. B. Ringelstein · S. Evers · M. Deppe
Department of Neurology, University of Münster,
Münster, Germany

S. S. Keller
Department of Clinical Neuroscience, Institute of Psychiatry,
King's College London,
London, UK

S. Mohammadi
Wellcome Trust Centre for Neuroimaging,
UCL Institute of Neurology,
University College London,
London, UK

C. Kellinghaus
Department of Neurology, Klinikum Osnabrück,
Osnabrück, Germany

H. Kugel · W. Schwindt
Department of Clinical Radiology, University of Münster,
Münster, Germany

K. Deppe
Department of Radiology, Klinikum Osnabrück,
Osnabrück, Germany

M. Deppe (✉)
Universität Münster, Klinik und Poliklinik für Neurologie,
Funktionelle Bildgebung,
Albert-Schweitzer-Campus 1, Gebäude A1,
48129 Münster, Germany
e-mail: mail@michael-deppe.de

Keywords FreeSurfer · Diffusion tensor imaging · Fractional anisotropy · MR image analysis · Stereology · Volume

Introduction

The thalamus is of central interest in many disorders of the nervous system (Andreasen 1997; Dom, et al. 1976; Lee and Marsden 1994; Meador-Woodruff, et al. 2003; Speedie and Heilman 1983; Williams 1965; Xuereb, et al. 1991). The functioning of the thalamus is crucial to many sensory, motor and cognitive systems, and therefore has also been subject to a great deal of investigation in the cognitive neurosciences (Basso, et al. 2005; Engelborghs, et al. 1998; Herrero, et al. 2002). It is in these capacities that

analysis of thalamic structure and function is a continually researched theme in neuroimaging investigations, particularly using magnetic resonance imaging (MRI). Analysis of volume or shape using MRI techniques may provide important information with respect to the involvement of the thalamus in neurological and neuropsychiatric disorders, including generalized (Du, et al. 2011; Pulsipher, et al. 2009) and partial (Gong, et al. 2008; Pulsipher, et al. 2007) epilepsy, schizophrenia (Adriano, et al. 2010), Huntington's disease (Douaud, et al. 2006; Kassubek, et al. 2005), Parkinson's disease (McKeown, et al. 2008), and Alzheimer's disease (de Jong, et al. 2008). Reliable measurement of thalamic structure is, however, notoriously difficult to achieve, particularly given the typically poor between-tissue MR contrast of the thalamic nuclei and adjacent white matter (Amini, et al. 2004). It is therefore important to develop new and improve and validate existing methodologies that provide thalamic metrics. Like for other subcortical brain structures, there are several approaches freely available to estimate thalamic volume from MR images. At either end of the MR image analysis spectrum, there are manual and automated approaches; manual approaches are user-dependent, time consuming but are considered to be the gold standard of MR image analysis techniques (Bonilha, et al. 2004; Collins and Pruessner 2010; Crum, et al. 2001; Pruessner, et al. 2000). Automated approaches remove the need of an expert anatomist, are dependent on computer algorithms, and are time efficient, but require a great deal of validation against manual methods to determine the specificity and validity of measurements (Chupin, et al. 2009; Morra, et al. 2008). The primary goal of the present study was to evaluate the validity of thalamic volume measurements obtained from a frequently used automated approach with respect to a reputable manual approach widely used in the imaging, anatomical and histological sciences.

The fully automated approach investigated in the present study was the subcortical segmentation and volume estimation techniques (Fischl, et al. 2002) incorporated into FreeSurfer software (<http://surfer.nmr.mgh.harvard.edu/>), which provide observer-independent volumes for individual subcortical nuclei from conventional MR images. Similarly to other methods that automatically segment and estimate subcortical volume such as FIRST (Patenaude, et al. 2011) incorporated into FSL software (<http://www.fmrib.ox.ac.uk/fsl/first/index.html>), there has been a recent proliferation of studies using FreeSurfer methods for volumetric studies, some of which have included comparison with manual methods, most notably for the hippocampus (Cherbuin, et al. 2009; Dewey, et al. 2010; Morey, et al. 2009; Pardoe, et al. 2009; Shen, et al. 2010; Tae, et al. 2008). To our knowledge, there has been no independent comparison between

manual methods and FreeSurfer methods for volume estimation of the thalamus. The manual approach used to evaluate FreeSurfer-based thalamic volumes in the present study was the Cavalieri method of design-based stereology in conjunction with point counting (Gundersen and Jensen 1987; Gundersen, et al. 1999; Mayhew 1992; Roberts, et al. 2000), which is a 100% investigator interactive technique that requires manual determination of sampling density for a given brain structure (i.e. the stereological parameters necessary to produce a reliable volume estimate) and investigator decisions on whether or not sampling probes (i.e. points) intersect the brain region-of-interest (ROI). Stereology requires the use of a human anatomist with expert knowledge of anatomical boundaries that divide legitimate (i.e. thalamic) and illegitimate (i.e. non-thalamic) brain tissue. Manual approaches such as stereology are considered gold standard because it is assumed that human knowledge and perception is superior to computer algorithms that determine regional brain boundaries.

We examined the consistency between manual and automated thalamic volume estimation in two ways. Firstly, we compared thalamic volume estimates in a sample of neurologically and psychiatrically healthy subjects. Secondly, we compared the methods in their sensitivity in detecting thalamic atrophy in patients with juvenile myoclonic epilepsy (JME). JME is an electro-clinical syndrome that by definition is non-lesional and without abnormality on conventional magnetic resonance imaging (MRI) (Berg, et al. 2010; ILAE 1989), but has previously been shown to be associated with thalamic structural alterations (Deppe, et al. 2008; Kim, et al. 2007; Mory, et al. 2011; Pulsipher, et al. 2009), and is generally considered to be intimately associated with thalamic dysfunction (Holmes, et al. 2010). We therefore compared morphometric approaches for volume estimation of both healthy and diseased thalami.

Methods

Participants

We studied a neurologically and psychiatrically healthy control group that was composed of 62 adult volunteers (32 females, mean age 27.9 ± 4.3 SD, range 21–43), all of whom had normal neurological examination and normal MRI (T1-, T2-weighted, and FLAIR). We also studied ten patients (6 females, mean age 28.6 ± 8.8 SD, range 19–42) with JME. Clinical information for these patients can be found elsewhere (Deppe, et al. 2008; Keller, et al. 2011a). There was no statistical difference in age between patients and controls ($t=0.38$, $p=0.70$). All subjects gave written informed consent and the local ethics committee approved this study.

Magnetic Resonance Imaging

All participants underwent high resolution MRI T1-weighted, T2-weighted and FLAIR imaging at 3 T (Philips Intera T30, T/R head coil). All MRI modalities were used to exclude the possibility of brain lesions in patients and controls. For volumetric analysis, we acquired T1-weighted structural MRIs using a high-resolution 3D turbo-field-echo sequence (matrix $256 \times 256 \times 160$ over a field of view of $25.6 \times 25.6 \times 16$ cm³ reconstructed after zero filling to $512 \times 512 \times 320$ cubic voxels with an edge length of 0.5 mm). Prior to morphometric analyses, all MR images were intensity inhomogeneity corrected and resampled to isotropic voxels of $1 \times 1 \times 1$ mm ($256 \times 256 \times 160$ slices) using in-house software (Eval 3.0). To confirm that there was no systematic difference in head size between patient and control groups, we automatically obtained relative brain size, VSCALE (global tissue volumes normalized for head size) and CSF volume estimates from the 3D T1-weighted images of all subjects using the SIENAX protocol (Smith, et al. 2002) integrated into FSL software (<http://www.fmrib.ox.ac.uk/fsl/siena/index.html>). There were no inter-group differences in relative brain size ($p=0.40$), VSCALE ($p=0.96$) or CSF volume ($p=0.95$).

A. Stereology

The Cavalieri method of design-based stereology in conjunction with point counting (Gundersen and Jensen 1987; Gundersen, et al. 1999; Mayhew 1992; Roberts, et al. 2000) was used as an unbiased estimator of the volume of the left and right thalamus in all subjects. By using the Cavalieri method, volume is directly estimated from equidistant and parallel MR images of the brain with a uniform random starting position. A second level of sampling is required to estimate the section area from each image by applying point counting within the ROI. The mathematical justification and implementation of the methodology is simple and it can be applied to structures of arbitrary shape (Garcia-Finana, et al. 2009). This technique has been frequently applied to reliably estimate brain volume and surface area on MR images (Acer, et al. 2010; Bas, et al. 2009; Cowell, et al. 2007; Eriksen, et al. 2010; Hallahan, et al. 2011; Howard, et al. 2003; Jelsing, et al. 2005; Keller, et al. 2009; Keller, et al. 2007; Keller, et al. 2002a; Keller, et al. 2009b; Keller, et al. 2002b; Lux, et al. 2008; Mackay, et al. 1998; Mackay, et al. 2000; Ronan, et al. 2006; Salmenpera, et al. 2005; Sheline, et al. 1996), and more widely applied to study other aspects of anatomy with and without the use of MRI. Stereology has been shown to be at least as precise as tracing and thresholding volumetry techniques and substantially more time efficient, with validation relative to post-mortem measurements (Garcia-Finana, et al. 2003; Garcia-Finana, et al. 2009; Keller and Roberts 2009; Keshavan, et al. 1995). Windows-

compatible Easymeasure software (Keller, et al. 2007; Puddephat 1999) was used for point counting on MR images.

Given that the borders of different thalamic nuclei are almost indistinguishable on conventional MRI, the thalamus was sampled as an entire complex, including the anterior thalamic nucleus, mediodorsal thalamic nucleus, lateral dorsal thalamic nucleus, ventral lateral thalamic nucleus, ventral postero-lateral/medial thalamic nuclei, and pulvinar (Duvernoy 1999). The lateral and medial geniculate nuclei were not included in measurements, similar to previous studies (Natsume, et al. 2003; Qiu, et al. 2009). On axial sections, point counting began on the dorsal most section, where the area of the lateral dorsal thalamic nucleus was demarcated laterally by the corona radiata and medially by the lateral ventricles. As measurements progressed ventrally, the posterior limb of the internal capsule bordered the thalamus laterally. On ventral sections, care was taken to delineate the thalamic nuclei from the neighbouring pars medialis of the globus pallidus, and at the most ventral sections, to segregate the pulvinar and area of the ventral posterolateral nucleus from the emerging hypothalamus anteriorly, superior colliculus posteromedially, and the hippocampus posteriorly / posterolaterally. Figure 1 shows point counting within the thalamic ROI relative to the automated FreeSurfer approach described below. More information on the MRI-based anatomy of the thalamus with respect to post-mortem sections can be found at <http://www.psychiatry.uiowa.edu/mhrc/pdf/papers/thalamus.pdf>.

The sampling density (i.e. size of points, distance between sections) were optimized to achieve a coefficient of error (CE) of less than 5% (Roberts, et al. 2000), an approach that has been adopted in our stereological analysis of the hippocampus (Keller, et al. 2002a; Keller, et al. 2002b), putamen (Keller, et al. 2011a), prefrontal region (Keller, et al. 2009a), planum temporale (Keller, et al. 2007; Keller, et al. 2011b), Broca's area (Keller, et al. 2007; Keller, et al. 2009b) and insula (Keller, et al. 2011b). The CE is essentially a statistical estimate of how accurate the stereological volume estimation is for each structure. Separation between test points on the square grid used for point counting was 0.312 cm (i.e., 4 pixels) and slice interval was 1 mm (singular axial MR sections). Thalamic transect area was obtained by multiplying the total number of points recorded by the area corresponding to each test point. An estimate of thalamic volume was obtained as the sum of the estimated areas of the structure transects on consecutive systematic sections multiplied by the distance between sections. Between approximately 400–550 points were recorded on approximately 20 systematic random sections.

B. FreeSurfer

FreeSurfer software (<http://surfer.nmr.mgh.harvard.edu/>) was used to obtain thalamic volumes for all subjects using

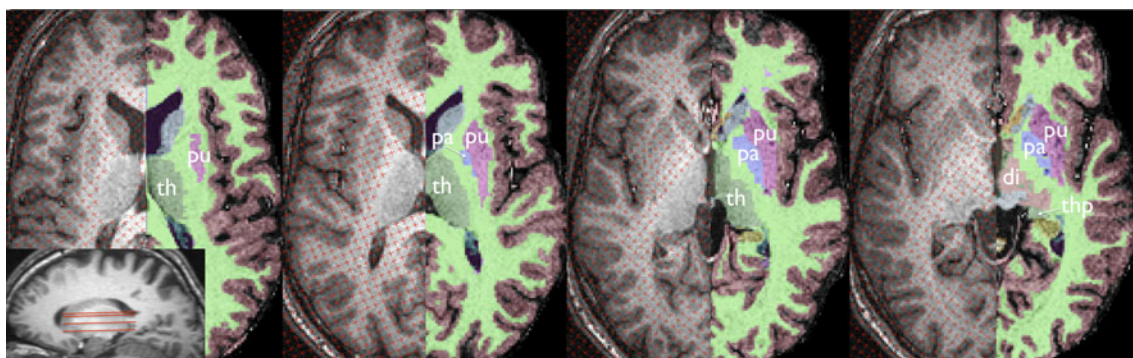


Fig. 1 Stereological and FreeSurfer methods used to estimate thalamic volume shown at approximately the same axial sections. Both methods are shown for the same hemisphere (the FreeSurfer axial sections are flipped to show maximal correspondence between techniques) of the same randomly selected subject. The sagittal MR section illustrates the

approximate levels of the axial sections shown. Abbreviations: di, ventral diencephalon (*peach*); pa, pallidum (*violet*); pu, putamen (*lilac*); th, thalamus (*dark green*); thp, pulvinar of the thalamus (*dark green*)

an observer-independent approach, which could be contrasted with the manual stereological measurements of the thalamus. Thalamic segmentations are based on the assignment of neuroanatomical labels to each voxel in an MR image based on the probabilistic information automatically estimated from a manually labelled training set. The methods of the automated volumetric approach have been described in detail previously (Fischl, et al. 2002), and the accuracy of automated labelling and volumetry of subcortical structures have been independently validated with respect to ‘gold standard’ manual volumetric techniques, predominantly for the hippocampus (Cherbuin, et al. 2009; Dewey, et al. 2010; Morey, et al. 2009; Pardoe, et al. 2009; Shen, et al. 2010; Tae, et al. 2008), and also of the amygdala (Dewey, et al. 2010; Morey, et al. 2009) and striatum (Dewey, et al. 2010). To our knowledge, there has been no independent comparison of the automated thalamic volumetry offered by FreeSurfer and a manual volumetric method (although thalamic tracings were compared with the performance of FreeSurfer in the original methods paper by Fischl et al. (2002)). Figure 1 shows the comparison of automated labelling of the thalamus (and extra-thalamic structures) in an individual control subject using FreeSurfer relative to stereological volume estimation of the thalamus in the same subject.

FreeSurfer analyses were performed on a Mac Pro (Version OS X 10.6.6, 32 GB, 2×2.93 GHz 6-Core Intel Xeon (HT)), which permitted the FreeSurfer ‘recon-all’ function (for cortical reconstruction and brain segmentation; <http://surfer.nmr.mgh.harvard.edu/fswiki/recon-all>) to complete 23 participants in less than 20 h. After the ‘recon-all’ function, the neuroanatomical labels were inspected for accuracy in all patients and controls. Despite that FreeSurfer permits manual editing to improve subcortical segmentation, no obvious errors in the automatic labelling were observed for any subject, and so all data obtained from FreeSurfer analyses were 100% automated and not influenced by manual intervention.

Statistical Analyses

Two-way mixed intra-class correlation coefficients for absolute agreement (Shrout and Fleiss 1979) were used to determine inter-rater agreement between two human raters using manual stereology and FreeSurfer for volume estimation of the thalamus in ten randomly selected controls using the statistics software SPSS (Version 18, www.spss.com). Intra-class correlations were subsequently performed between stereological volumes obtained by one human rater and FreeSurfer volumes for the entire sample of patients and

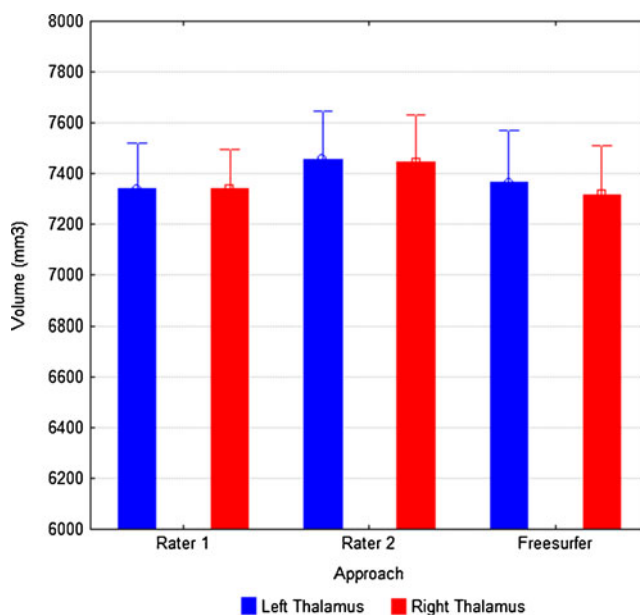


Fig. 2 Inter-rater consistency in stereological volume estimation of the left (*blue*) and right (*red*) thalamus, and relation to volume estimates obtained from FreeSurfer. Stereological volume estimates are reproducible, and FreeSurfer is entirely consistent with the volumes obtained with both rater one and rater two. Error bars indicate the 95% confidence intervals

Table 1 Inter-rater intra-class coefficients for volumetric measures

	Rater 1	Rater 2	FS
Left thalamus			
Rater 1	–	0.977	0.969
Rater 2	0.977	–	0.925
FS	0.969	0.925	–
Right thalamus			
Rater 1	–	0.952	0.923
Rater 2	0.952	–	0.916
FS	0.923	0.916	–

controls ($n=72$). Univariate ANOVAs were used to investigate patient-control differences in volumes, and corrected for multiple comparisons using Statistica version 9.1, (Stat Soft. Inc, www.statsoft.com).

Results

A. Stereology vs FreeSurfer: Consistency between volumetric measures

Figure 2 shows the comparison of the three approaches (rater one (R1) for stereology, rater two (R2) for stereology and FreeSurfer) to estimate thalamic volume in the randomly selected ten subjects. This inter-rater / between-technique analysis indicates consistency across measures, and most notably, that FreeSurfer performed at least as consistent as R2 relative to R1. Mean (SD) left and right thalamic volume was 7339.6 mm^3 (567.3) and 7339.2 mm^3 (489.3) for R1, 7456.3 mm^3 (597.3) and 7444.3 mm^3 (590.6) for R2, and 7365.0 mm^3 (641.3) and 7317.6 mm^3 (610.0) for FreeSurfer, respectively. Intra-class correlations across raters and

approaches are presented in Table 1. Measurements between R1 and R2, and between manual raters and FreeSurfer, achieve high intra-class correlations (all >0.9).

Figure 3 presents the relationship between stereological (R1) and FreeSurfer estimates of left and right thalamus volume across the entire sample of 72 subjects investigated in the present study. It is immediately obvious that the two methods yield consistent thalamic volumes. Intra-class correlations revealed a slightly reduced level of consistency between human and FreeSurfer analysis of the entire sample relative to the sub-sample of ten subjects (left thalamus= 0.812 , right thalamus= 0.881). Mean (SD) left and right thalamic volume was 7422.4 mm^3 (824.3) and 7390.1 mm^3 (805.8) for stereology and 7343.6 mm^3 (824.3) and 7388.3 mm^3 (844.0) for FreeSurfer. There was no difference between the volume of the left and right thalamus across all 72 subjects using stereology ($F=0.09$, $p=0.82$) or FreeSurfer ($F=0.10$, $p=0.75$) (and no differences when patients and controls were separated, $p>0.80$). Although there were occasional differences in the direction of inter-hemispheric thalamic volume asymmetry in individual cases between stereology and FreeSurfer (Fig. 3, right panel), this did not represent a statistically significant group effect ($F=1.46$, $p=0.23$).

B. Stereology vs FreeSurfer: Identification of thalamic atrophy in JME

Both techniques were equally sensitive in detecting bilateral thalamic volume atrophy in patients with JME relative to controls (Fig. 4). Using stereology, mean (SD) left and right thalamic volume was 6843.2 mm^3 (746.6) and 6763.3 mm^3 (824.0) in patients with JME, and 7507.8 mm^3 (805.6) and 7482.6 mm^3 (767.2) in controls, respectively. Volume reduction in patients was found to be statistically significant for the left ($F(1,70)=5.43$, $p=0.02$) and right ($F(1,70)=6.77$, $p=0.01$) thalamus compared to controls. Using FreeSurfer, mean (SD) left and right thalamic volume was 6803.4 mm^3 (732.8) and 6750.0 mm^3 (714.3) in patients with JME, and 7430.7 mm^3 (809.9) and 7491.3 mm^3 (822.3) in controls, respectively.

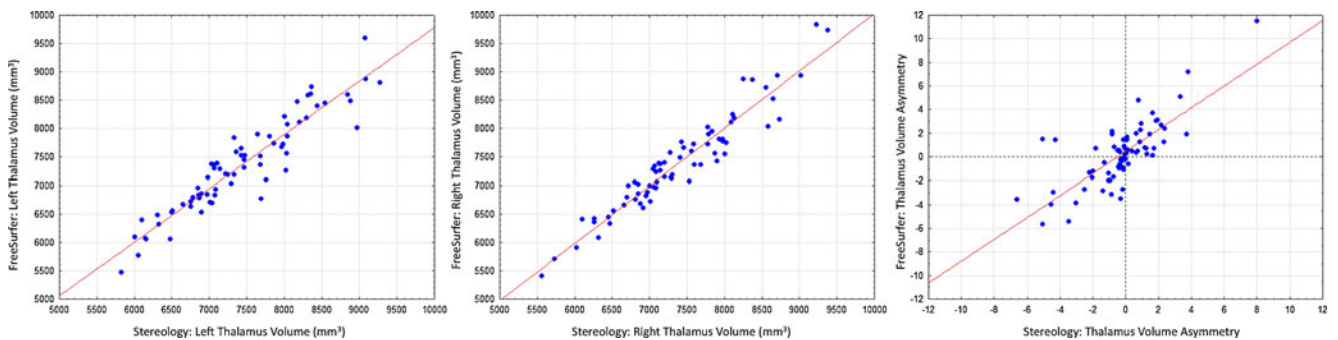


Fig. 3 Relationship between stereological and FreeSurfer volume estimates of the left **a** and right **b** thalamus in the whole study sample. The relationship between left-right asymmetries determined for each individual participant by each technique is shown in **c**. Although stereology and

FreeSurfer determined left-right asymmetries of the thalamus in the same direction for the vast majority of subjects (lower left and upper right quadrants), there were some disassociations (*top left quadrant*)

FreeSurfer thalamic volumes were similarly smaller in patients relative to controls in the left ($F(1,70)=4.76, p=0.03$) and right ($F(1,70)=7.36, p=0.008$) hemispheres.

Discussion

The volume of the thalamus is a notoriously difficult metric to estimate reliably given the low contrast between thalamic gray matter and adjacent white matter on T1-weighted MR images, which is a particular challenge for automated MR image analysis methods (Amini, et al. 2004). Only by comparing such automated methods with manual investigator-intensive methods can we establish the reliability of volume estimates. The present study provides important data indicating the specificity and validity of automated thalamic volume estimation using FreeSurfer software. In particular, further to demonstrating consistency between stereological and FreeSurfer volume estimates of the thalamus in healthy subjects and neurological patients, we demonstrate that the extent of agreeability between stereology and FreeSurfer is equal to the agreeability between two human anatomists estimating thalamic volume using stereological methods.

FreeSurfer software is now a frequently used tool for the estimation of subcortical structure volume. At the time of writing, a pubmed search using “Freesurfer” and “volume” yields 87 articles (October 2011). The vast majority of these articles are application studies, particularly in neurological disorders, and only a few have sought to evaluate the validity of volume measurements. Various levels of consistency between FreeSurfer and manual ROI methods have been reported for the hippocampus (Cherbuin, et al. 2009; Dewey, et al. 2010; Morey, et al. 2009; Pardoe, et al. 2009; Shen, et al. 2010; Tae, et al. 2008), amygdala (Dewey, et al. 2010; Morey, et al. 2009) and striatum (Dewey, et al. 2010). Dewey et al. (2010) performed a series of comparisons between the fully automated techniques of FreeSurfer and Individual Brain Atlases using Statistical Parametric Mapping (IBASPM) with auto-assisted manual tracings of the hippocampus, amygdala, putamen and caudate. The authors report that FreeSurfer segmentations exhibited significantly higher mean spatial overlap with auto-assisted tracings in all structures compared to IBASPM using dice coefficients. We were not in a position to perform spatial overlap analyses of the thalamus given that stereology and FreeSurfer are two inherently distinct MR image analysis approaches. However, this is one of the primary strengths of the data presented here, inasmuch that a reliable volume estimate obtained using a gold-standard (non-voxel labelling) manual approach on MR images without automated spatial transformations (i.e. in native space) is comparable to a fully automated approach that requires spatial transformations in order to label an ROI and obtain

a volume. Our interest was with respect to the reliability of the volume estimate of the thalamus.

To our knowledge, the present study is the first to independently provide data validating the application of FreeSurfer to obtain automated volumes of the left and right thalamus. Based on the congruence between the data obtained from FreeSurfer and manual stereology—the latter of which is considered to represent the ‘gold standard’ approach due to the requirement of an expert anatomist—we recommend the use of FreeSurfer software for accurate volumetric quantification of the thalamus using high-resolution T1-weighted MRI. The removal of an expert

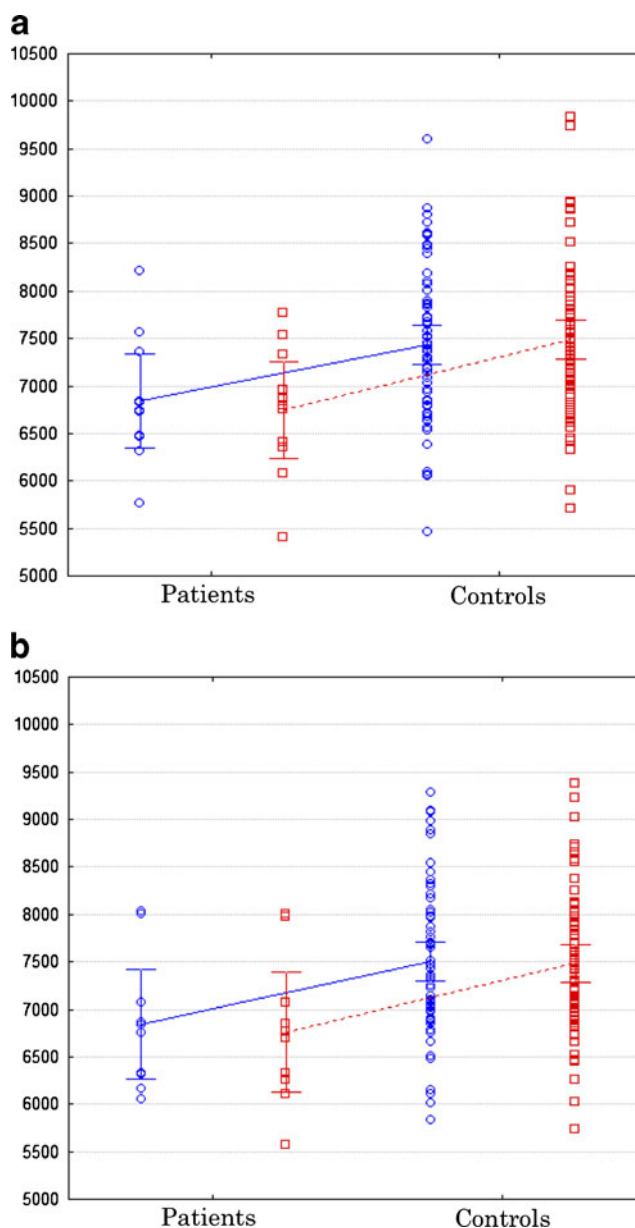


Fig. 4 Volume reduction of the left (blue circles) and right (red squares) thalamus in patients with JME relative to healthy controls using stereology **a** and FreeSurfer **b**. Error bars indicate the 95% confidence intervals

anatomist for volumetric analyses is cost effective and time efficient, particularly in large-scale volumetric studies. Importantly, we demonstrate that the automated technique is as sensitive in detecting pathological alterations of the thalamus relative to stereology, which promotes the use of FreeSurfer in neurological contexts.

There are two additional issues that should be highlighted. Measurements made in the present study were of global thalamic volume. The thalamus is composed of lamellae that segregate multiple nuclei with distinct connections and functions, which are likely to be differentially affected in various neurological and neuropsychiatric disorders. For example, in disorders where the thalamus is implicated in patients also exhibiting deficits in frontal lobe functioning—such as in JME (Pulsipher, et al. 2009)—it would be expected that anterior thalamic nuclei that project to the frontal lobe would be preferentially affected (Deppe, et al. 2008). In such circumstances it will be interesting to investigate structural alterations of differential thalamic subregions, which are measures that the techniques applied in the present study cannot provide. There are other techniques that may provide the basis for quantitative measurements of thalamic subregions based on DTI and quantitative T1 and T2 imaging (Behrens, et al. 2003; Johansen-Berg, et al. 2005; Traynor, et al. 2011). Secondly, the global estimates of thalamic volume using FreeSurfer in the present study was obtained on a Philips Intera T30 3 T MRI system, requiring no additional manual edits for (obvious) incorrect labelling of the thalamic ROI after the application of our in-house image inhomogeneity and resampling algorithm. Different MRI systems and head coils may have different image contrast characteristics that can potentially affect the performance of automated MR image analysis techniques. However, reproducibility of FreeSurfer estimated thalamic volume from serially acquired MR images on the same MR system is high (Jovicich, et al. 2009; Morey, et al. 2010), and MR system manufacturer has been shown to have little effect on volume estimates (Jovicich, et al. 2009).

In summary, this study provides convincing evidence for the reliability of global thalamic measurements using FreeSurfer in healthy and damaged thalami. The use of this software is cost effective and particularly advantageous in large-scale cross-sectional studies and longitudinal investigations in neurological settings.

Information Sharing Statement

FreeSurfer software is publicly and freely available from the FreeSurferWiki resource (<http://surfer.nmr.mgh.harvard.edu/fswiki/FreeSurferWiki>), which is developed and maintained at the Martinos Center for Biomedical Imaging (<http://www.nmr.mgh.harvard.edu/martinos/noFlashHome.php>). All

software, information and support are provided online at the FreeSurferWiki webpage. Easymeasure software for volume estimation using stereology is freely available from the authors of this manuscript upon request. Dr. Mike Puddephat developed Easymeasure software at the University of Liverpool, UK. Further information can be found at (<http://www.easymeasure.co.uk/>).

Acknowledgements This work was supported by the Transregional Collaborative Research Centre SFB/TR 3 (Project A8) of the Deutsche Forschungsgemeinschaft (DFG). EBR acknowledges support from the Neuromedical Foundation (Stiftung Neuromedizin), Münster. SM acknowledges support from the Wellcome Trust; Open access to the paper were supported by the Wellcome Trust grant number 091593/Z/10/Z.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Acer, N., Cankaya, M. N., Isci, O., Bas, O., Camurdanoglu, M., & Turgut, M. (2010). Estimation of cerebral surface area using vertical sectioning and magnetic resonance imaging: a stereological study. *Brain Research*, *1310*, 29–36.
- Adriano, F., Spoletini, I., Caltagirone, C., & Spalletta, G. (2010). Updated meta-analyses reveal thalamus volume reduction in patients with first-episode and chronic schizophrenia. *Schizophrenia Research*, *123*(1), 1–14.
- Amini, L., Soltanian-Zadeh, H., Lucas, C., & Gity, M. (2004). Automatic segmentation of thalamus from brain MRI integrating fuzzy clustering and dynamic contours. *IEEE Transactions on Biomedical Engineering*, *51*(5), 800–811.
- Andreasen, N. C. (1997). The role of the thalamus in schizophrenia. *Canadian Journal of Psychiatry*, *42*(1), 27–33.
- Bas, O., Acer, N., Mas, N., Karabekir, H. S., Kusbeci, O. Y., & Sahin, B. (2009). Stereological evaluation of the volume and volume fraction of intracranial structures in magnetic resonance images of patients with Alzheimer's disease. *Annals of Anatomy*, *191*(2), 186–195.
- Basso, M. A., Uhlrich, D., & Bickford, M. E. (2005). Cortical function: a view from the thalamus. *Neuron*, *45*(4), 485–488.
- Behrens, T. E., Johansen-Berg, H., Woolrich, M. W., Smith, S. M., Wheeler-Kingshott, C. A., Boulby, P. A., Barker, G. J., Sillery, E. L., Sheehan, K., Ciccarelli, O., Thompson, A. J., Brady, J. M., & Matthews, P. M. (2003). Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nature Neuroscience*, *6*(7), 750–757.
- Berg, A. T., Berkovic, S. F., Brodie, M. J., Buchhalter, J., Cross, J. H., van Emde, B. W., Engel, J., French, J., Glauser, T. A., Mathern, G. W., Moshe, S. L., Nordli, D., Plouin, P., & Scheffer, I. E. (2010). Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005–2009. *Epilepsia*, *51*(4), 676–685.
- Bonilha, L., Kobayashi, E., Cendes, F., & Min, Li L. (2004). Protocol for volumetric segmentation of medial temporal structures using high-resolution 3-D magnetic resonance imaging. *Human Brain Mapping*, *22*(2), 145–154.
- Cherbuin, N., Anstey, K. J., Reglade-Meslin, C., & Sachdev, P. S. (2009). In vivo hippocampal measurement and memory: a

- comparison of manual tracing and automated segmentation in a large community-based sample. *PLoS One*, 4(4), e5265.
- Chupin, M., Hammers, A., Liu, R. S., Colliot, O., Burdett, J., Bardinet, E., Duncan, J. S., Garnero, L., & Lemieux, L. (2009). Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: method and validation. *NeuroImage*, 46(3), 749–761.
- Collins, D. L., & Pruessner, J. C. (2010). Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. *NeuroImage*, 52(4), 1355–1366.
- Cowell, P. E., Sluming, V. A., Wilkinson, I. D., Cezayirli, E., Romanowski, C. A., Webb, J. A., Keller, S. S., Mayes, A., & Roberts, N. (2007). Effects of sex and age on regional prefrontal brain volume in two human cohorts. *European Journal of Neuroscience*, 25(1), 307–318.
- Crum, W. R., Scahill, R. I., & Fox, N. C. (2001). Automated hippocampal segmentation by regional fluid registration of serial MRI: validation and application in Alzheimer's disease. *NeuroImage*, 13(5), 847–855.
- de Jong, L. W., van der Hiele, K., Veer, I. M., Houwing, J. J., Westendorp, R. G., Bollen, E. L., de Bruin, P. W., Middelkoop, H. A., van Buchem, M. A., & van der Grond, J. (2008). Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. *Brain*, 131(Pt 12), 3277–3285.
- Deppe, M., Kellinghaus, C., Duning, T., Moddel, G., Mohammadi, S., Deppe, K., Schifflbauer, H., Kugel, H., Keller, S. S., Ringelstein, E. B., & Knecht, S. (2008). Nerve fiber impairment of anterior thalamocortical circuitry in juvenile myoclonic epilepsy. *Neurology*, 71(24), 1981–1985.
- Dewey, J., Hana, G., Russell, T., Price, J., McCaffrey, D., Harezlak, J., Sem, E., Anyanwu, J. C., Guttmann, C. R., Navia, B., Cohen, R., & Tate, D. F. (2010). Reliability and validity of MRI-based automated volumetry software relative to auto-assisted manual measurement of subcortical structures in HIV-infected patients from a multisite study. *NeuroImage*, 51(4), 1334–1344.
- Dom, R., Malfroid, M., & Baro, F. (1976). Neuropathology of Huntington's chorea. Studies of the ventrobasal complex of the thalamus. *Neurology*, 26(1), 64–68.
- Douaud, G., Gaura, V., Ribeiro, M. J., Lethimonnier, F., Maroy, R., Verny, C., Krystkowiak, P., Damier, P., Bachoud-Levi, A. C., Hantraye, P., & Remy, P. (2006). Distribution of grey matter atrophy in Huntington's disease patients: a combined ROI-based and voxel-based morphometric study. *NeuroImage*, 32(4), 1562–1575.
- Du, H., Zhang, Y., Xie, B., Wu, N., Wu, G., Wang, J., Jiang, T., & Feng, H. (2011). Regional atrophy of the basal ganglia and thalamus in idiopathic generalized epilepsy. *Journal of Magnetic Resonance Imaging*, 33(4), 817–821.
- Duvernoy, H. (1999). *The human brain. Surface, blood supply and three-dimensional sectional anatomy*. New York: New York Springer.
- Engelborghs, S., Marien, P., Martin, J. J., & De Deyn, P. P. (1998). Functional anatomy, vascularisation and pathology of the human thalamus. *Acta Neurologica Belgica*, 98(3), 252–265.
- Eriksen, N., Rostrup, E., Andersen, K., Lauritzen, M. J., Fabricius, M., Larsen, V. A., Dreier, J. P., Strong, A. J., Hartings, J. A., & Pakkenberg, B. (2010). Application of stereological estimates in patients with severe head injuries using CT and MR scanning images. *British Journal of Radiology*, 83(988), 307–317.
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., & Dale, A. M. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3), 341–355.
- Garcia-Finana, M., Cruz-Orive, L. M., Mackay, C. E., Pakkenberg, B., & Roberts, N. (2003). Comparison of MR imaging against physical sectioning to estimate the volume of human cerebral compartments. *NeuroImage*, 18(2), 505–516.
- Garcia-Finana, M., Keller, S. S., & Roberts, N. (2009). Confidence intervals for the volume of brain structures in Cavalieri sampling with local errors. *Journal of Neuroscience Methods*, 179(1), 71–77.
- Gong, G., Concha, L., Beaulieu, C., & Gross, D. W. (2008). Thalamic diffusion and volumetry in temporal lobe epilepsy with and without mesial temporal sclerosis. *Epilepsy Research*, 80(2–3), 184–193.
- Gundersen, H. J., & Jensen, E. B. (1987). The efficiency of systematic sampling in stereology and its prediction. *Journal of Microscopy*, 147(Pt 3), 229–263.
- Gundersen, H. J., Jensen, E. B., Kieu, K., & Nielsen, J. (1999). The efficiency of systematic sampling in stereology—reconsidered. *Journal of Microscopy*, 193(Pt 3), 199–211.
- Hallahan, B. P., Craig, M. C., Toal, F., Daly, E. M., Moore, C. J., Ambikapathy, A., Robertson, D., Murphy, K. C., & Murphy, D. G. (2011). In vivo brain anatomy of adult males with Fragile X syndrome: an MRI study. *NeuroImage*, 54(1), 16–24.
- Herrero, M. T., Barcia, C., & Navarro, J. M. (2002). Functional anatomy of thalamus and basal ganglia. *Child's Nervous System*, 18(8), 386–404.
- Holmes, M. D., Quiring, J., & Tucker, D. M. (2010). Evidence that juvenile myoclonic epilepsy is a disorder of frontotemporal corticothalamic networks. *NeuroImage*, 49(1), 80–93.
- Howard, M. A., Roberts, N., Garcia-Finana, M., & Cowell, P. E. (2003). Volume estimation of prefrontal cortical subfields using MRI and stereology. *Brain Research. Brain Research Protocols*, 10(3), 125–138.
- ILAE. (1989). Proposal for revised classification of epilepsies and epileptic syndromes. Commission on Classification and Terminology of the International League Against Epilepsy. *Epilepsia*, 30(4), 389–399.
- Jelsing, J., Rostrup, E., Markenroth, K., Paulson, O. B., Gundersen, H. J., Hemmingsen, R., & Pakkenberg, B. (2005). Assessment of in vivo MR imaging compared to physical sections in vitro—a quantitative study of brain volumes using stereology. *NeuroImage*, 26(1), 57–65.
- Johansen-Berg, H., Behrens, T. E., Sillery, E., Ciccarelli, O., Thompson, A. J., Smith, S. M., & Matthews, P. M. (2005). Functional-anatomical validation and individual variation of diffusion tractography-based segmentation of the human thalamus. *Cerebral Cortex*, 15(1), 31–39.
- Jovicich, J., Czanner, S., Han, X., Salat, D., van der Kouwe, A., Quinn, B., Pacheco, J., Albert, M., Killiany, R., Blacker, D., Maguire, P., Rosas, D., Makris, N., Gollub, R., Dale, A., Dickerson, B. C., & Fischl, B. (2009). MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: Reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. *NeuroImage*, 46(1), 177–192.
- Kassubek, J., Juengling, F. D., Ecker, D., & Landwehrmeyer, G. B. (2005). Thalamic atrophy in Huntington's disease co-varies with cognitive performance: a morphometric MRI analysis. *Cerebral Cortex*, 15(6), 846–853.
- Keller SS, Ahrens T, Mohammadi S, Moddel G, Kugel H, Bernd Ringelstein E, Deppe M. (2011a). Microstructural and volumetric abnormalities of the putamen in juvenile myoclonic epilepsy. *Epilepsia* In press.
- Keller, S. S., Baker, G., Downes, J. J., & Roberts, N. (2009). Quantitative MRI of the prefrontal cortex and executive function in patients with temporal lobe epilepsy. *Epilepsy & Behavior*, 15(2), 186–195.
- Keller, S. S., Highley, J. R., Garcia-Finana, M., Sluming, V., Rezaie, R., & Roberts, N. (2007). Sulcal variability, stereological

- measurement and asymmetry of Broca's area on MR images. *Journal of Anatomy*, 211(4), 534–555.
- Keller, S. S., Mackay, C. E., Barrick, T. R., Wieshmann, U. C., Howard, M. A., & Roberts, N. (2002). Voxel-based morphometric comparison of hippocampal and extrahippocampal abnormalities in patients with left and right hippocampal atrophy. *NeuroImage*, 16(1), 23–31.
- Keller, S. S., & Roberts, N. (2009). Measurement of brain volume using MRI: software, techniques, choices and prerequisites. *J Anthropological Sci*, 87, 127–151.
- Keller, S. S., Roberts, N., Garcia-Finana, M., Mohammadi, S., Ringelstein, E. B., Knecht, S., & Deppe, M. (2011). Can the Language-dominant Hemisphere Be Predicted by Brain Anatomy? *Journal of Cognitive Neuroscience*, 23(8), 2013–2029.
- Keller, S. S., Roberts, N., & Hopkins, W. (2009). A Comparative Magnetic Resonance Imaging Study of the Anatomy, Variability, and Asymmetry of Broca's Area in the Human and Chimpanzee Brain. *Journal of Neuroscience*, 29(46), 14607–14616.
- Keller, S. S., Wieshmann, U. C., Mackay, C. E., Denby, C. E., Webb, J., & Roberts, N. (2002). Voxel based morphometry of grey matter abnormalities in patients with medically intractable temporal lobe epilepsy: effects of side of seizure onset and epilepsy duration. *Journal of Neurology, Neurosurgery, and Psychiatry*, 73(6), 648–655.
- Keshavan, M. S., Anderson, S., Beckwith, C., Nash, K., Pettegrew, J. W., & Krishnan, K. R. (1995). A comparison of stereology and segmentation techniques for volumetric measurements of lateral ventricles in magnetic resonance imaging. *Psychiatry Research*, 61(1), 53–60.
- Kim, J. H., Lee, J. K., Koh, S. B., Lee, S. A., Lee, J. M., Kim, S. I., & Kang, J. K. (2007). Regional grey matter abnormalities in juvenile myoclonic epilepsy: a voxel-based morphometry study. *NeuroImage*, 37(4), 1132–1137.
- Lee, M. S., & Marsden, C. D. (1994). Movement disorders following lesions of the thalamus or subthalamic region. *Movement Disorders*, 9(5), 493–507.
- Lux, S., Keller, S., Mackay, C., Ebers, G., Marshall, J. C., Cherkas, L., Rezaie, R., Roberts, N., Fink, G. R., & Gurd, J. M. (2008). Crossed cerebral lateralization for verbal and visuo-spatial function in a pair of handedness discordant monozygotic twins: MRI and fMRI brain imaging. *Journal of Anatomy*, 212(3), 235–248.
- Mackay, C. E., Roberts, N., Mayes, A. R., Downes, J. J., Foster, J. K., & Mann, D. (1998). An exploratory study of the relationship between face recognition memory and the volume of medial temporal lobe structures in healthy young males. *Behavioural Neurology*, 11(1), 3–20.
- Mackay, C. E., Webb, J. A., Eldridge, P. R., Chadwick, D. W., Whitehouse, G. H., & Roberts, N. (2000). Quantitative magnetic resonance imaging in consecutive patients evaluated for surgical treatment of temporal lobe epilepsy. *Magnetic Resonance Imaging*, 18(10), 1187–1199.
- Mayhew, T. M. (1992). A review of recent advances in stereology for quantifying neural structure. *Journal of Neurocytology*, 21(5), 313–328.
- McKeown, M. J., Uthama, A., Abugharbieh, R., Palmer, S., Lewis, M., & Huang, X. (2008). Shape (but not volume) changes in the thalami in Parkinson disease. *BMC Neurology*, 8, 8.
- Meador-Woodruff, J. H., Clinton, S. M., Beneyto, M., & McCullumsmith, R. E. (2003). Molecular abnormalities of the glutamate synapse in the thalamus in schizophrenia. *Annals of the New York Academy of Sciences*, 1003, 75–93.
- Morey, R. A., Petty, C. M., Xu, Y., Hayes, J. P., Wagner, H. R., 2nd, Lewis, D. V., LaBar, K. S., Styner, M., & McCarthy, G. (2009). A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *NeuroImage*, 45(3), 855–866.
- Morey, R. A., Selgrade, E. S., Wagner, H. R., 2nd, Huettel, S. A., Wang, L., & McCarthy, G. (2010). Scan-rescan reliability of subcortical brain volumes derived from automated segmentation. *Human Brain Mapping*, 31(11), 1751–1762.
- Morra, J. H., Tu, Z., Apostolova, L. G., Green, A. E., Avedissian, C., Madsen, S. K., Parikshak, N., Hua, X., Toga, A. W., Jack, C. R., Jr., Weiner, M. W., & Thompson, P. M. (2008). Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls. *NeuroImage*, 43(1), 59–68.
- Mory, S. B., Betting, L. E., Fernandes, P. T., Lopes-Cendes, I., Guerreiro, M. M., Guerreiro, C. A., Cendes, F., & Li, L. M. (2011). Structural abnormalities of the thalamus in juvenile myoclonic epilepsy. *Epilepsy & Behavior*, 21(4), 407–411.
- Natsume, J., Bernasconi, N., Andermann, F., & Bernasconi, A. (2003). MRI volumetry of the thalamus in temporal, extratemporal, and idiopathic generalized epilepsy. *Neurology*, 60(8), 1296–1300.
- Pardoe, H. R., Pell, G. S., Abbott, D. F., & Jackson, G. D. (2009). Hippocampal volume assessment in temporal lobe epilepsy: How good is automated segmentation? *Epilepsia*, 50(12), 2586–2592.
- Patenaude B, Smith S, Kennedy D, Jenkinson M. (2011). A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage*.
- Pruessner, J. C., Li, L. M., Serles, W., Pruessner, M., Collins, D. L., Kabani, N., Lupien, S., & Evans, A. C. (2000). Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: minimizing the discrepancies between laboratories. *Cerebral Cortex*, 10(4), 433–442.
- Puddephat, M. (1999). *Computer interface for convenient application of stereological methods for unbiased estimation of volume and surface area: studies using MRI with particular reference to the human brain. The Magnetic Resonance and Image Analysis Research Centre (MARIARC)*. Liverpool: University of Liverpool.
- Pulsipher, D. T., Seidenberg, M., Guidotti, L., Tuchscherer, V. N., Morton, J., Sheth, R. D., & Hermann, B. (2009). Thalamofrontal circuitry and executive dysfunction in recent-onset juvenile myoclonic epilepsy. *Epilepsia*, 50(5), 1210–1219.
- Pulsipher, D. T., Seidenberg, M., Morton, J. J., Geary, E., Parrish, J., & Hermann, B. (2007). MRI volume loss of subcortical structures in unilateral temporal lobe epilepsy. *Epilepsy & Behavior*, 11(3), 442–449.
- Qiu, A., Zhong, J., Graham, S., Chia, M. Y., & Sim, K. (2009). Combined analyses of thalamic volume, shape and white matter integrity in first-episode schizophrenia. *NeuroImage*, 47(4), 1163–1171.
- Roberts, N., Puddephat, M. J., & McNulty, V. (2000). The benefit of stereology for quantitative radiology. *British Journal of Radiology*, 73(871), 679–697.
- Ronan, L., Doherty, C. P., Delanty, N., Thornton, J., & Fitzsimons, M. (2006). Quantitative MRI: a reliable protocol for measurement of cerebral gyrfication using stereology. *Magnetic Resonance Imaging*, 24(3), 265–272.
- Salmenpera, T., Kononen, M., Roberts, N., Vanninen, R., Pitkanen, A., & Kalviainen, R. (2005). Hippocampal damage in newly diagnosed focal epilepsy: a prospective MRI study. *Neurology*, 64(1), 62–68.
- Sheline, Y. I., Black, K. J., Lin, D. Y., Christensen, G. E., Gado, M. H., Brunsten, B. S., & Vannier, M. W. (1996). Stereological MRI volumetry of the frontal lobe. *Psychiatry Research*, 67(3), 203–214.
- Shen, L., Saykin, A. J., Kim, S., Firpi, H. A., West, J. D., Risacher, S. L., McDonald, B. C., McHugh, T. L., Wishart, H. A., & Flashman, L. A. (2010). Comparison of manual and automated determination of hippocampal volumes in MCI and early AD. *Brain Imaging and Behavior*, 4(1), 86–95.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428.

- Smith, S. M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P. M., Federico, A., & De Stefano, N. (2002). Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *NeuroImage*, *17*(1), 479–489.
- Speedie, L. J., & Heilman, K. M. (1983). Anterograde memory deficits for visuospatial material after infarction of the right thalamus. *Archives of Neurology*, *40*(3), 183–186.
- Tae, W. S., Kim, S. S., Lee, K. U., Nam, E. C., & Kim, K. W. (2008). Validation of hippocampal volumes measured using a manual method and two automated methods (FreeSurfer and IBASPM) in chronic major depressive disorder. *Neuroradiology*, *50*(7), 569–581.
- Traynor, C. R., Barker, G. J., Crum, W. R., Williams, S. C., & Richardson, M. P. (2011). Segmentation of the thalamus in MRI based on T1 and T2. *NeuroImage*, *56*(3), 939–950.
- Williams, D. (1965). The thalamus and epilepsy. *Brain*, *88*(3), 539–556.
- Xuereb, J. H., Perry, R. H., Candy, J. M., Perry, E. K., Marshall, E., & Bonham, J. R. (1991). Nerve cell loss in the thalamus in Alzheimer's disease and Parkinson's disease. *Brain*, *114*(Pt 3), 1363–1379.