Tina Memo No. 2006-004
Short version published in Proc. MIUA 2006
Long version published in Medical Image Analysis (http://dx.doi.org/10.1016/j.media.2008.10.006)

A Comparative Evaluation of Cortical Thickness Measurement Techniques

P. A. Bromiley, M.L.J. Scott, and N.A. Thacker

Last updated
16 / 6 / 2006

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

Author's note: The material included in this memo was subsequently published in a paper in Medical Image Analysis: a preprint is available as TINA Memo No. 2009-001.

A Comparative Evaluation of Cortical Thickness Measurement Techniques

P. A. Bromiley, M.L.J. Scott, and N.A. Thacker.
Imaging Science and Biomedical Engineering Division
Medical School, University of Manchester
Manchester, M13 9PT, UK
paul.bromiley@man.ac.uk

Abstract

In vivo measurements of cortical thickness from MR images have potentially widespread utility in the characterisation of normal brain development and maturation as well as in diagnosing and measuring the progress of a number of cortical pathologies. The literature describes several approaches to this problem, which may be divided into two groups: those relying on deformable models of the inner and outer cortical surfaces, and those relying on image intensities alone. Results from the former may be largely model driven at points deep within sulci, where no apparent channel of cerebrospinal fluid can be seen at the resolution of typical MR images, potentially introducing bias. We present a comparative evaluation of cortical thickness measurement techniques, which demonstrates that approaches based on edge detection can provide cortical thickness measurements of equal accuracy to model-based approaches, using less processor time, and without the possibility of bias from a model.

1 Introduction

The human cerebral cortex makes up the largest part of the brain, and consists of a highly convoluted layer of neuronal cells with the topology of a 2D sheet, surrounding a core of white matter. Its thickness varies considerably, from approximately 2mm in the calcarine sulcus to approximately 4mm in the precentral gyrus, with an average of approximately 3mm [1, 6, 9, 10]. Measurements of cortical thickness have shown considerable potential both in the study of normal brain growth and maturation [16, 28, 26, 21, 32] and in the diagnosis, or measurement of the progression, of a wide variety of cortical pathologies including Alzheimer's disease [30], Huntingdon's disease [20], schizophrenia [12, 30], Williams syndrome [31], attention-deficit-hyperactivity disorder [32], and fetal alcohol syndrome [32]. Reliable and automated techniques for *in vivo* cortical thickness measurement therefore form a useful tool in neurology.

Any cortical thickness measurement technique requires two core components: a method for locating the inner and outer cortical surfaces, and a metric with which to measure the distance between them. A variety of thickness metrics have been suggested, varying from simply measuring the thickness along normals to the inner cortical surface [15] to approaches based on partial differential equations [9, 27, 14]. However, the thickness metric chosen must be considered the definition of the quantity to be measured: the accuracy with which this measurement is made is dictated by the method chosen to define the cortical surfaces. The techniques presented in the literature can be coarsely divided into two groups in this respect: model-based (top-down) approaches that involve fitting deformable models to the inner and outer cortical surfaces i.e. the boundaries between white matter (WM) and grey matter (GM) and GM and pia matter (PM), and data-driven (bottom-up) approaches that detect these interfaces using image intensities alone. The deformable model based approach typically involves segmenting the WM, fitting a model to the WM/GM interface, and then expanding this surface until it reaches the GM/PM interface. The inner cortical boundary is smoother, and so fitting this boundary first helps to guide the model into the more complex folding of the outer boundary. The model itself typically consists of a set of vertices which tessellate the cortical surface with triangular elements. Early versions of this approach [4, 7, 20] used three terms in the objective function used to expand the surface: a spring term based on the distances between vertices, which enforced an approximately even coverage of vertices over the surface, a curvature term, which enforced local smoothness of the model, and an intensity term, which located the GM/PM interface. However, it is also desirable to prevent self-intersections in the models, ensuring that they have a simple, spherical topology, which in turn allows the surface and subsequent thickness measurements to be mapped onto a plane or sphere using standard projections. This can be advantageous both in terms of displaying the results and, more importantly, in the definitions of the distance metrics used (see below). Therfore, later algorithms introduced topological constraints to prevent

self-intersections, either by explicitly removing them when they occurred [6] or by introducing additional terms into the objective function. The latter approach is used in the ASP (anatomic segmentation using proximities) algorithm [15], which introduces both a surface self-proximity term and a term based on the distance between corresponding vertices on the inner and outer surface models. Other refinements of the basic approach hae also been attempted: for example, the CRUISE (Cortical Reconstruction Using Implicit Surface Evolution) algorithm ([33] and references therein) introduces an additional model surface in the centre of the cortical ribbon, and the approach described in [16] involves fitting this central surface alone, measuring the distance from each vertex to the inner cortical boundary, and then doubling this value to find an approximation to the cortical thickness.

The deformable model based approach has two main drawbacks. First, such algorithms require considerable computational resources, largely due to the topological constraints. For example, the algorithms presented in [15] and [6] required 30hrs on a 180 MHz Silicon Graphics R10000 and 5 hours on a 500 MHz Pentium 3 respectively to process each data set. Second, the introduction of terms to prevent self-intersections of the model surfaces may bias the algorithm towards a fixed separation between the inner and outer cortical surface models [15], depending on the weights assigned to each term in the objective function. Such terms may be required in order to solve the notoriously difficult problem of fitting the model in tightly folded gyri [11, 31], where the sulcal banks oppose so closely that there is no clear CSF channel in the sulcus at the resolution of typical MR images. In such regions the outer surface model may fail to fit the pial surface within the sulcal fundus, leading to thickness estimates that are at least two times too high. Such biases may therefore be inevitable. These problems have led to an interest in data-driven techniques, in which the inner and outer cortical surfaces are determined using only image intensities (e.g. [13, 31, 14]). Such approaches also typically begin with segmentation of the WM, GM and cerebrospinal fluid (CSF). Thickness measurements are then performed at each point on the inner cortical surface, by propagating away from the surface according to some thickness metric until the outer cortical surface is reached. One notable advantage of such approaches is that, in regions of tightly folded gyri as described above, if the pial surface is missed and another point on the inner cortical surface is reached, the thickness measurement can be halved to produce a value which is approximate but still entirely data-driven and so free from model bias. It has been shown that this approach has little effect on final, regional thickness measurements if local smoothing is applied [31, 14]. Some authors have also attempted to enhance the CSF in deep sulci, skeletonise the result and subtract this from the original image to "scour" the sulci, forcing gaps to appear at points where the opposing sulcal banks appear to touch and thus making detection of the outer cortical surface easier [9]. Alternatively, morphological operators can be used to perform the same task.

The second requirement for cortical thickness measurement is a metric with which to measure the thickness itself. In the case of algorithms such as ASP, this can be based directly on the vertices of the inner and outer cortical models themselves, either by taking the distance between homologous vertices, finding the closest outer surface vertex to each inner surface vertex, or finding the distance to the outer surface along a local normal to the inner surface [15]. Similar approaches can be adopted in data-driven techniques, finding the local surface normal at each point on the inner cortical surface and then propagating to the pial surface [13]. However, more complex approaches have also been attempted. For example, the inner and outer cortical surfaces can be treated as charged conductors and the Laplace equation applied to define hypothetical electric field lines between them [9]. The field lines mimic the known neuroanatomy of the cortex, which consists of a set of distinct layers [5]. This approach has also been applied to the ASP algorithm [11]. Other algorithms such as marching cubes or the Eikonal fire equation [22] have also been applied [27, 14]. Some authors have also combined the two approaches by fitting deformable models in order to define homologous points on the cortical surfaces across different subjects and using these in registration, for defining the points at which thickness measurements are made, and to define a surface on which the results may be displayed, whilst still using only image intensities in the thickness measurement itself [27, 14], although this approach might still introduce model bias. Finally, some authors have avoided the problems of measuring cortical thickness altogether by measuring GM density instead i.e. the proportion of GM within a kernel around each point on the cortical surface: this quantity is closely associated with cortical thickness [26].

In previous work we have presented a data-driven cortical thickness measurement technique [24]. In this paper, we compare the results from this algorithm to a wide range of published measurements, produced using both model-based and data-driven algorithms. The aim was to determine whether the use of deformable models to define the inner and outer cortical boundaries increases the accuracy of the thickness measurement (i.e. reduces the random errors) or introduces bias (i.e. systematic errors).

2 Method

The cortical thickness measurement technique used in this study has been presented previously [24]; we summarise the method here. Several stages of preprocessing were used. Initially, the partial volume segmentation technique

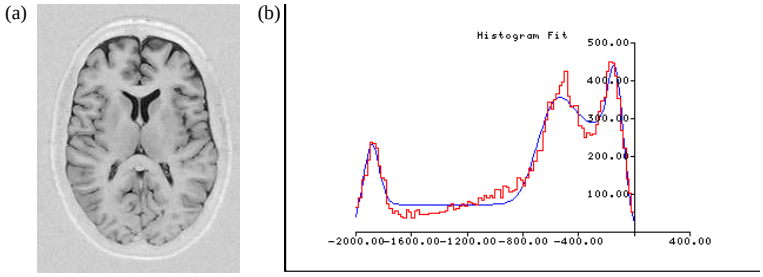


Figure 1: Example axial inversion-recovery image (a) and corresponding grey-level intensity histogram (b) showing, from left to right, peaks for the CSF, GM and WM.

described in [18], which involves fitting a Bayesian mixture model containing both pure tissue and partial volume terms to the image histogram (as shown in fig. 1), was applied to the GM, WM and CSF. This process produced measurements of the means and standard deviations of the pure tissue grey levels and measurements of the most likely tissue volume contributions in each voxel. To obtain a finer through-plane resolution whilst preserving tissue boundaries, the data was explicitly up-interpolated [17] using a partial volume scheme to constrain the potential tissue boundaries, determined using 3D image gradients, that could pass through a partial volume voxel. Finally, a map of the GM was produced. In addition, the original image data was registered to a common stereotaxic space (the Talairach atlas [29]) using a linear affine transform. The atlas defined the 31 cortical regions used later in producing regional histograms of the cortical thickness: its use allowed direct comparison to results presented in [10].

The cortical thickness was measured using a modified edge detection process (Canny [3]) to determine the GM/WM boundary. A 'z-score' measure of the grey-level of each voxel being consistent with the GM/WM midpoint value was used to construct a likelihood image which highlighted the GM/WM boundary. This was then used as a replacement for the conventionally used sum-squared image gradient (edge strength) map in the Canny edge detector in order to produce well localised connected edge strings to sub-pixel accuracy. The 3D surface normal at each voxel on the GM/WM boundary was determined by taking the local grey-level gradient of the 3D, Gaussian-smoothed (using a kernel of [1/2, 5]), up-interpolated grey-level data. The GM tissue probability maps and the GM/WM boundary and edge orientations were used to determine the distance from the boundary, at each voxel on the boundary, along the orientation direction to another GM edge (see [23] for details). This edge could either be a GM/CSF boundary or, if no CSF was visible in the intervening sulcus, a GM/WM boundary, and was determined by comparing the value of the voxel in the original grey-level image to the mean GM value. If it was a WM boundary it was assumed that the opposing banks of a sulcus had been traversed and the sulcal thickness was taken to be half this length: this approach has been shown to have little effect on subsequent regional thickness measurements [31, 14]. Histograms of the cortical thickness measurements in each region were then produced and their median values taken to produce the final regional thickness measurements. The regional histograms typically contained between 100 and 6000 entries, thus giving reasonable anatomical precision as well as a robust estimate of the median.

The method was applied to inversion-recovery MR scans (1.5T Philips ACS PT 6000 NT, TI/TR/TE = 300/6850/18 ms, pixel size = 0.9x0.9mm, 51 slices) of 119 normal volunteers (52 male, mean age = 70.3 years, range = 19-86 years). 110 scans had axial slices (thickness = 3.0mm), 9 had coronal (thickness = 4.0mm). The results were then compared to cortical thickness measurements from the literature in two stages. First, regional average thickness measurements in a range of standard cortical regions were computed and compared to the results published by Kabani et al. [10], which were produced using a deformable model-based technique. Only the youngest 13 subjects (mean age 36.9 years, range 19-53) were included, to ensure that the mean age of the group was similar to that used by Kabani et al. Linear fits to cortical thickness vs. age in all 110 subjects were used to correct for the remaining age difference. This analysis acted as an exemplar comparison between model-based and data-driven techniques. Second, a meta-study of precentral gyrus thickness variation with age was performed, involving measurements from nine studies including this one, and incorporating 635 subjects. This region is the location of the primary motor cortex, and is also the thickest region of the cortex; however, the choice of region was solely based on the number of previously published thickness measurements available in the literature. It is known that changes in both cortical thickness [27] and overall brain volume [2] occur with age. Any bias introduced by the use of deformable models would suppress the ability to observe such changes, and so this meta-study provides a more detailed evaluation of such biases.

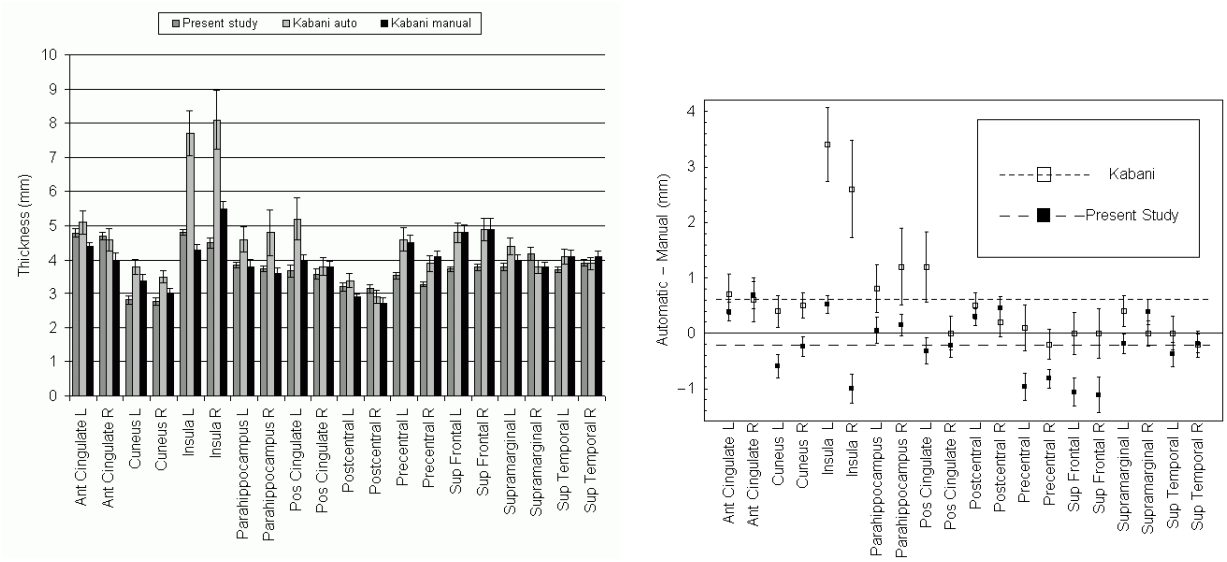
3 Results

Kabani et al. [10] presented regional average cortical thickness measurements in ten regions from 40 subjects produced using the ASP algorithm. Half of the subjects were used in each hemisphere, giving an effective group size of 20. In addition, they presented manual measurements of the same quantities. Figure 2a shows a comparison of these results to thickness measurements in the same regions produced using the algorithm presented here. Assuming that the manual measurements provide a gold standard, Fig. 2b provides a more quantitative comparison, showing the difference between the manual measurements and the results from each algorithm. The mean differences across all regions are 0.61 ± 0.43 mm for the ASP algorithm and -0.21 ± 0.22 mm for the algorithm presented here. Neither difference is statistically significant, and so there is no evidence of systematic error in either algorithm. Given that the group sizes are similar in each case (20 in Kabani et al. vs. 13 in the present study) the statistical errors from the algorithm presented here are approximately half as large as those on the Kabani et al. results.

Figure 3 shows the average cortical thickness measurements for the precentral gyrus in 119 subjects produced using the algorithm presented here, plotted against age. A quadratic fit to the data is shown: the dashed curves either side of the fit show the upper and lower standard error bounds. A significant ($P < 0.0001$) reduction of cortical thickness with age is observed. In previous work [25] we found similar dependencies in other cortical regions. Also shown are a number of measurements of precentral gyrus thickness from the literature. The details of the studies involved are given in Table 1. With the exception of the results from Kabani et. al. [10], the data were read from graphical representations[1]. The results presented by von Economo [5] were measured manually post-mortem: brain volume decreases by approximately 10% during postmortem fixation [19]. However, the thickness measurement was performed only on the gyral cap, which is known to be thicker than the sulcal fundi [15]. Similarly the presentation of the results in Sowell et. al. [27] and Thompson et. al. [31] as projections onto the outer cortical surface prevented identification of the thickness in the sulcal fundi. These three data therefore represent upper limits on the average thickness in the region. Overall, the studies represented in Fig. 3 represent the widest possible range of methods for defining the inner and outer cortical surfaces, the thickness metric, and the presentation of the results. Some variation between the measurements, introduced by these differences in experimental procedure, might therefore be expected. However, the data show a remarkable level of agreement both with each other and the results from this study. If the errors on the results quoted by Kabani et. al. [10] can be taken as representative of the errors on the other studies, then there is no statistically significant difference between any of these results and our own.

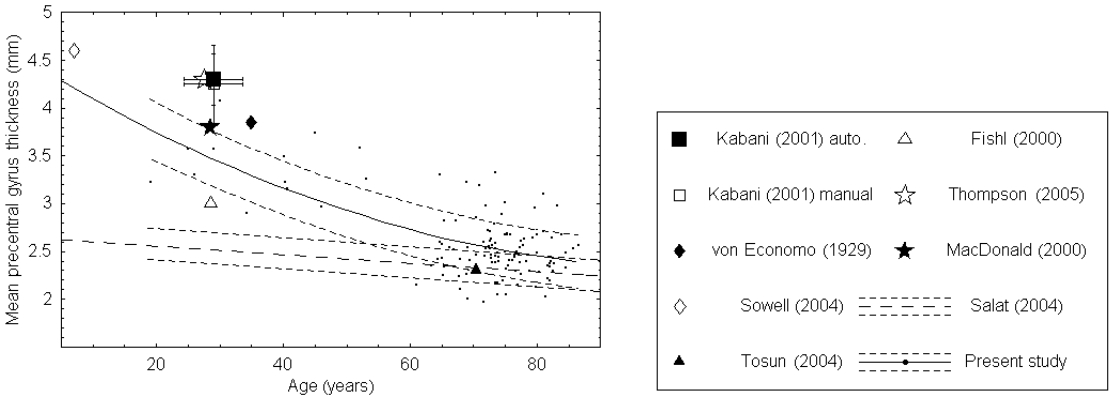
The remaining paper included in this study, Salat et al. [21], was the only one to study variation in cortical thickness with age, and the only model-based study to cover the whole age range between adolescence and senescence. A significant disagreement between this study and the others included in the meta-study can be seen, with Salat et al. suggesting a much lower rate of cortical thickness change with age. Given the level of agreement between the other studies, and the fact that the algorithm used by Salat et al. was based on deformable modelling of the inner and outer cortical surfaces, we suggest that this may be due to the possibility for bias identified by [15] in such algorithms. The constraints applied to prevent self-intersection of the models and to aid modelling of the surfaces in tightly folded gyri tend to bias the model towards a fixed thickness, suppressing the observation of age-related changes. The same effect was found in the mean rate of thickness change over the whole cortex: a global rate of 0.016 ± 0.0052 mm year^{-1} was found in the present study, which is consistent with the 0.0077 $mm\ year^{-1}$ found in [16], but is an order of magnitude higher than, and inconsistent with, the 0.0016 mm year^{-1} rate quoted by Salat et al. Finally, age-related changes in GM density i.e. the proportion of GM within a kernel around each point on the cortical surface, a quantity closely related to cortical thickness, were measured in [26]: the average, proportional rate of change of GM density in the precentral gyrus was 0.593 % $year^{-1}$, consistent with the 0.567 ± 0.270 % $year^{-1}$ cortical thickness change found in the present study.

[1] With the exception the results from Salat et. al. [21] and von Economo [5] these were presented as views of the outer cortical surface, in some cases partially inflated to reveal the sulcal fundi, with colour coding to represent the thickness at each point. This method of data display is popular in the literature as it avoids the need for parcellation of the data into particular regions. However, the calculation of regional average thicknesses from such representations is difficult and the calculation of errors on the averages impossible. Hence, these points are shown without error bars.



(a) (b)

Figure 2: Manual and automatic regional average cortical thickness measurements presented in [10], and automatic results from the algorithm presented here (a), and differences between the manual thickness measurements and the algorithm results (b); the dashed lines show average differences across all regions for each algorithm.



(a) (b)

Figure 3: Measurements of the average cortical thickness in the precentral gyrus: see main text for description.

Reference	No. subjects	Age range (years)	Algorithm type
Kabani et. al. 2001 [10]	40	18-40	Model based
von Economo 1929 [5]	-	30-40	Manual measurement
Sowell et. al. 2004 [27]	45	5-11	Intensity based
Tosun et. al. 2004 [33]	105	59-84	Model based
Fishl et. al. 2000 [6]	30	20-37	Model based
Thompson et. al. 2005 [31]	40	18-48	Intensity based
MacDonald et. al. 2000 [15]	150	18-40	Model based
Salat et. al. 2004 [21]	106	18-93	Model based

Table 1: Details of the studies included in the meta-study of the dependence of precentral gyrus thickness on age.

4 Conclusion

The literature describes several approaches to cortical thickness measurement, many of which rely deformable models fitted to the inner and outer cortical surfaces. The topological constraints applied in such algorithms require considerable amounts of processor time, typically several hours per image volume. In addition, bias may be introduced by the model at points where the information available from the image is weak. In previous work we presented a simpler alternative, based on edge detection followed by measurement of the thickness along normals to the inner cortical surface, which requires an order of magnitude less processor time and does not introduce bias. We have presented comparisons of the results from this algorithm against a range of measurements from the literature. The comparisons indicate that results from this algorithm are more accurate than those published in the literature, suggesting that the accuracy of both types of algorithm is dictated by the initial segmentation. However, in at least one study, the use of an algorithm based on deformable models appears to have suppressed the ability to detect age-related change. We therefore conclude that the use of deformable models in cortical thickness measurement provides no advantage in terms of accuracy.

The comparison of age-related changes in cortical thickness can also be used to investigate the ability of the algorithm presented here to detect pathological effects. Comparison with previous results shows that the thickness measurements in the precentral gyrus are consistent with those presented in the literature over a wide age range, confirming the accuracy of the technique. In previous work [24] we have shown that statistically significant age-related changes were detected in this data set throughout the cortex. Previous studies on age-related and pathological brain volume change [8] have shown that pathological changes typically occur an order of magnitude faster than age-related changes. Therefore, the ability to detect age-related changes strongly implies that pathological changes will also be detected.

Acknowledgements

This work was supported by the University of Manchester Translational Imaging Unit and funded by the MIAS (Medical Images and Signals) IRC under EPSRC grant no. GR/N14248/01 and the UK Medical Research Council grant no. D2025/31, and by the Wellcome Trust grant no. ME003889. The software used in this study is freely available from www.tina-vision.net.

References

[1] K Brodmann. *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prenzipien dargestellt auf Grund des Zellaufbaus*. Springer, Barth, Leipzig, Germany, 1909.

[2] P A Bromiley, N A Thacker, and A Jackson. Trends in brain volume change with normal ageing. Proc. MIUA, 2005.

[3] J F Canny. A computational approach to edge detection. *IEEE PAMI*, 8(6):679–698, Nov 1986.

[4] A M Dale, B Fischl, and M I Sereno. Cortical surface-based analysis 1. segmentation and surface reconstruction. *NeuroImage*, 9:179–194, 1999.

[5] C Von Economo. *The Cytoarchitectonics of the Human Cerebral Cortex*. Trasnlasted by S. Parker. Oxford University Press, Oxford, 1929.

[6] B Fischl and A M Dale. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *PNAS*, 97:11050–11055, 2000.

[7] B Fischl, M I Sereno, and A M Dale. Cortical surface-based analysis. *NeuroImage*, 9:195–207, 1999.

[8] N C Fox, S Cousens, R Scahill, R J Harvey, and M N Rossor. Using serial registered brain magnetic resonance imaging to measure disease progression in Alzheimer's disease. *Archives of Neurology*, 57:339–344, 2000.

[9] S E Jones, B R Buchbinder, and I Aharon. Three-dimensional mapping of cortical thickness using Laplace's equation. *Human Brain Mapping*, 11:12–32, 2000.

[10] N Kabani, G Le Goualher, D MacDonald, and A C Evans. Measurement of cortical thickness using an automated 3-D algorithm: A validation study. *NeuroImage*, 13:375–380, 2001.

[11] J S Kim, V Singh, J K Lee, J Lerch, Y Ad-Dab'bagh, D MacDonald, J M Lee, S I Kim, and A C Evans. Automated 3-d extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification. *NeuroImage*, 27:210–221, 2005.

[12] G R Kuperberg, M R Broome, P K McGuire, A S David, M Eddy, F Ozawa, D Goff, W C West, S C R Williwma, A J W van der Kouwe, D H Salat, A M Dale, and B Fischl. Regionally localised thinning of the cerebral cortex in schizophrenia. *Arch Gen Psychiatry*, 60:878–888, 2006.

[13] G Lohmann, C Preul, and M Hund-Georgiadis. Morphology-based cortical thickness estimation. In Christopher J. Taylor and J. Alison Noble, editors, *Information Processing in Medical Imaging, 18th International Conference, IPMI 2003, Ambleside, UK, July 20-25, 2003, Proceedings*, volume 2732 of *Lecture Notes in Computer Science*, pages 89–100. Springer, 2003.

[14] E Luders, K L Narr, P M Thompson, D E Rex, L Janke, and A W Toga. Hemispheric asymmetries in cortical thickness. *Cerebral Cortex*, page In Press, 2006.

[15] D MacDonald, N Kabani, D Avis, and A C Evans. Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *NeuroImage*, 12:340–356, 2000.

[16] V A Magnotta, N C Andreasen, S K Schultz, G Harris, T Cizadlo, D Heckel, P Nopoulos, and M Flaum. Quantitative in vivo measurement of gyrification in the human brain: Change associated with aging. *Cerebral Cortex*, 9:151–160, 1999.

[17] S McKie and N A Thacker. Step interpolation of MR images with inter-slice gap correction. Tina Memo 2003-010, www.tina-vision.net/memos/2003-010.pdf, 2003.

[18] M Pokrić, N Thacker, M L J Scott, and A Jackson. Multi-dimensional medical image segmentation with partial voluming. In *Proc. MIUA*, pages 77–81, 2001.

[19] R Quester and R Schroder. The shrinkage of the human brain stem during formalin fixation and embedding in paraffin. *J. Neurosci. Methods*, 75:81–89, 1997.

[20] H D Rosas, A K Liu, S Hersch, M Glessner, R J Ferrante, D H Salat, A van der Kouwe, B G Jenkins, A M Dale, and B Fischl. Regional and progressive thinning of the cortical ribbon in huntingdon's disease. *Neurology*, 58:695–701, 2002.

[21] D H Salat, R L Buckner, A Z Snyder, D N Greve, R S R Desikan, E Busa, J C Morris, A M Dale, and B Fischl. Thinning of the cerebral cortex in aging. *Cerebral Cortex*, 14:721–730, 2004.

[22] G Sapiro. *Geometric partial differential equations and image analysis*. Cambridge University Press, Cambridge, 2001.

[23] M L J Scott and N A Thacker. Cerebral cortical thickness measurements. Tina memo 2004-007, www.tina-vision.net/docs/2004-007.pdf, 2004.

[24] M L J Scott and N A Thacker. Regional cerebral cortical thickness measurement using MRI. In *Proc. MIUA*, pages 7–10, 2005.

[25] M L J Scott and N A Thacker. Robust tissue boundary detection for cerebral cortical thickness estimation. Tina memo 2005-007, www.tina-vision.net/docs/2005-007.pdf, 2005.

[26] E R Sowell, B S Peterson, P M Thompson, S E Welcome, A L Henkenius, and A W Toga. Mapping cortical change across the human life span. *Nature Neuroscience*, 6:309–315, 2003.

[27] E R Sowell, P M Thompson, C M Leonard, S E Welcome, E Kan, and A W Toga. Longitudinal mapping of cortical thickness and brain growth in normal children. *J. Neurosci.*, 24:8223–8231, 2004.

[28] E R Sowell, P M Thompson, K D Tessner, and A W Toga. Mapping continued brain growth and grey matter density reduction in dorsal frontal cortex: Inverse relationships during postadolescent brain maturation. *J. Neurosci*, 21:8819–8829, 2001.

[29] J Talairach and P Tournoux. *Co-planner stereotaxic atlas of the human brain. 3-Dimensional proportional system: An approach to cerebral imaging*. Thieme Medical Publishers, New York, 1988.

[30] P M Thompson, K M Hayashi, E R Sowell, N Gogtay, J N Geidd, J l Rapoport, G I de Zubicarary, A L Janke, S E Rose, J Semple, D M Moddrell, Y Wang, T G M van Erp, T D Cannon, and A W Toga. Mapping cortical change in Alzheimer's disease, brain development, and schizophrenia. *NeuroImage*, 23:S2–S18, 2004.

[31] P M Thompson, A D Lee, R A Dutton, J A Geaga, K M Hayashi, M A Eckert, U Bellugi, A M Galaburda, J R Korenberg, D L Mills, A W Toga, and A L Reiss. Abnormal cortical complexity and thickness profiles mapped in Williams syndrome. *J. Neurosci.*, 25:4146–4158, 2005.

[32] A W Toga, P M Thompson, and E R Sowell. Mapping brain maturation. *Trends in Neurosciences*, page In Press, 2006.

[33] D Tosun, M E Rettmann, X Han, X Tao, C Xu, S M Resnick, D L Pham, and J L Prince. Cortical surface segmentation and mapping. *NeuroImage*, 23:S108–S118, 2004.

5 Appendix: MIUA 2006 Reviewer's Comments

A slightly shortened version of this paper was submitted to MIUA 2006. The reviewer's comments we received seem to indicate that, in some cases, the main message of the paper was misunderstood. Therefore, the reviewer's comments are given below (italicised) together with our reponses (in normal text), in the hope that they may be useful in understanding the work.

6 Review 1

The paper compares a previously published intensity-based algorithm for measuring cortical thickness from MRI scans with model-based techniques in the literature. The hypothesis is that the faster intensity-based method is less biased. The paper is well written and clear. The results of using the model-based methods used in the comparison are taken from various publications in the literature rather than from the same data set from which results of the intensity-based method come. This is a significant weakness of the paper, as questions remain about whether the results are directly comparable and no real test of bias has been performed. The conclusions are conjecture.

The development of a measurement technique for any physiological variable is predicated upon two implicit yet important assumptions: that the quantity being measured has a well-defined mean, and that the normal variation around that mean is small compared to the changes introduced by pathological processes. If either of these assumptions are not met, then there is no point in measuring that variable, since it can never be used to characterise a biological process. In the case of cortical thickness, many papers have been published demonstrating correlations with diseases: references are given in the introduction to our paper. Therefore, we can conclude that the two assumptions are met in the case of cortical thickness measurement. In that case, all cortical thickness measurement techniques should give comparable measurements when the primary confounding variable of age is taken into account and the studies are performed on normals from the same demographics, as were the studies included in our paper. Indeed, this concept of repeatability is a central feature of the scientific method, and is the reason why meta studies are so extensively used in clinical research.

It is true that little, beyond the need for a third study, can be concluded from the comparison of two studies if they disagree. However, when multiple studies are available we can build up a detailed picture of the performance of the various measurement techniques used. In our paper we performed two comparisons. First, we compared cortical thickness measurements in a number of regions generated using our intensity-based algorithm to the numerical results presented by Kabani et. al. (2001), the most detailed set of measurements available in the literature, which were generated using the model-based ASP algorithm. We showed that there was a negligible difference between the two studies in terms of random errors. In the second comparison we focused on a single region, and compared all available measurements from the literature. We demonstrated that the vast majority of studies provided results consistent with our own. However, the results from the highly cited Salat et al. (2004) paper, the only study to measure the dependence of cortical thickness measurements on age over a wide range using a model-based technique, disagreed with all of the other studies included in the meta-study. Salat et al. give a rate of change with age that is much lower than that implied by all of the other studies. MacDonald et al, (2000), in their original publication of the ASP algorithm used by Salat et al., pointed out the potential for bias introduced by the use of deformable models, and showed that it would bias cortical thickness measurements towards a fixed value i.e. suppress, for example, any age-related changes. In our paper, we invoked Occam's razor, using the MacDonald paper to explain the Salat et al. results.

This left us with two conclusions: a) that our intensity-based technique was in general no less accurate than the model-based techniques (based on the first comparison) and b) that bias due to the use of deformable models was implicated in at least one of the other studies (based on the second comparison). Whilst the second could certainly be called a hypothesis, it can hardly be called conjecture.

7 Review 2

I have no argument with the authors' method or the fact that it is likely to be superior to analysis using deformable models; however, the method has been published previously and it is on the comparative evaluation that this paper has to be judged.

Meta-analysis is a useful tool for comparing different studies usually comprising large cohorts, e.g. in the justification of radiation protection. For smaller samples it has its faults, as the studies are based on different groups introducing local bias, the influence of confounding variables, etc. It may still be the way to go if there is no other option; however, in this case, it appears to have been used as the easy way out.

Our meta-study incorporated 635 subjects: approaching the number that would be included in a phase 3 clinical trial (typically 1000-3000+). By what definition could this be described as a "small" sample? For example, could the reviewer tell us how many other papers presented at MIUA this year describe studies with larger samples?

I have added an explicit statement of the number of subjects included in the meta-study to the paper.

Surely, algorithms such as ASP are readily available and, if not, could be programmed by the authors or their programming colleagues. I can see no reason why the different approaches could not be applied to the authors'

own sample of images of normal subjects. This would have greatly strengthened the conclusions that are reached. To base these conclusions on the poor results of one study, which may have been flawed for reasons other than methodology, is less than convincing, particularly when these results are displayed (Figure 2a) without error bars.

We strongly disagree with the principle behind this statement, on two grounds. First, replicating the software of another group is far more problematic than the reviewer implies. If the software is reimplemented, then any discrepancy in the results can always be dismissed as the result of bugs in the code. Even if the original software is obtained from the group in question, there are always control parameters in such algorithms, which can be implicated in any variation in the results. However, second and more important is the fact that all of the measurements included in the meta-study were taken from published papers, the authors of which stated in every case that they were producing accurate estimates of the cortical thickness. If that statement is true, and with reference to my reply to the first reviewer, then the results should be comparable if the sample size is large: in this case, with 635 subjects included, it is eminently large enough to allow comparison. If that statement is not true, then I suggest the reviewer write to each journal concerned to ask for the papers to be withdrawn.

The lack of error bars on Figure 2a is extensively discussed in the paper, and arises for a variety of reasons. In some cases, the results quoted are expected to be upper or lower limits on the mean cortical thickness. In most cases, the prevailing technique of presenting cortical thickness measurements as colourmaps projected onto the outer cortical surface prevents all but order-of-magnitude estimates of the errors being made. However, I have added error bars to the Salat et al. results by reconstituting the data from the graph in the original paper and repeating the fitting process. The conclusions however are not changed: the Salat et al results are significantly different from the rest of the studies at younger ages.

We maintain that our conclusions are valid, and entirely consistent with application of the scientific method. MacDonald et al. (2000) predicted the possibility of bias in the ASP algorithm, which would manifest itself as a bias towards a single thickness, suppressing any age-related changes. Salat et al. have published results which significantly disagree with the rest of the literature in exactly the way predicted. Invoking Occam's razor, we hypothesise that (we use the phrase "appears that" in the conclusions to make it clear that this conclusion is based on inference rather than definitive proof) model bias has led to this effect. Of course, this hypothesis may be overturned by subsequent investigation. However, we can also rely on the doctrine of falsifiability: a single model-based cortical thickness study that demonstrates bias-like effects is enough to call the method into question, which is all we are doing in this paper. It should also be noted that, as stated in the paper, the Salat et al. study is the only model-based study to cover the full age range from adolescence to senescence.

The authors state that the mean rate of thickness change found by the authors (0.016 +/- 0.0052 mm/year) is consistent with that found elsewhere (0.0077 mm/year)), as opposed to that found by Salat et al (0.0016 mm/year). I would have thought that there is considerable inconsistency between all three estimates.

Salat et al. (2004) does not quote errors on his rate measurement: Magnotta et al. (1999) provides graphical results (from which this rate was calculated) but bins his subjects into four coarse age ranges, making it impossible to produce an error estimate. However, visual inspection of the graphs of results in both papers show that the distribution about the mean cortical thickness is either equivalent to (in the case of Magnotta) or narrower than (in the case of Salat) the distribution observed in our own results. Therefore, let us make the conservative assumption that the results as accurate as our own. We can then use the error estimate from our own rate measurement on the results from the other papers.

Comparing the Magnotta et. al. result to our own: Difference: $0.016 - 0.0077 = 0.0083$ Error: $Sqrt(0.0052^2 + 0.0052^2) = 0.0074$ i.e. the results vary by a little more than 1 s.d.: not significant

Comparing the Salat et al. result to our own: Difference: $0.016 - 0.0016 = 0.0144$ Error: $Sqrt(0.0052^2 + 0.0052^2) = 0.0074$ i.e. the results vary by 2 s.d.: significant at the 95% confidence limit

I have slightly modified the relevant section of the paper to make the statement clearer.

I am also less than impressed by the authors' suggestion to halve the thickness value when it is realised that their method has got it wrong.

"...got it wrong" is a trivialisation of the description of the method, and indicates that the reviewer has missed the whole point of the paper. Partial voluming is an inevitable feature in MR images: the CSF channel may be completely obscured by this effect at points deep within thin sulci. Therefore, any cortical thickness measurement technique has to deal with situations where the pial surface cannot be seen. Model-based methods attempt to deal with this using constraints on the model, which force the outer surface model to follow the inner surface model into deep sulci: this has the potential to introduce bias, since at these points there is no information available from the data to locate the outer cortical surface, and the fitting of the model is thus entirely driven by the constraints. The purpose of our earlier work on the development of the cortical thickness algorithm was to investigate the extent to which the measurement could be made without the use of deformable models, thus eliminating this potential

source of bias. However, the problem of locating the outer cortical surface must still be dealt with. We chose to take a data-driven approach: at points where the outer cortical surface cannot be located, searches along the local normal to the inner cortical surface will traverse the sulcus and find the inner cortical surface in the opposing sulcal bank. In the absence of any data to specify the thicknesses of the opposing banks in this situation, we make the most conservative assumption possible by assuming equal thicknesses. Whilst this assumption is known to be inaccurate at certain points in the cortex (e.g. the anterior and posterior banks of the central sulcus) it is the only non-model-based approach available. Previous authors (e.g. Thompson et al. 2005, Luders et al. 2006, both of which are cited in our paper) have demonstrated that this approach has minimal effects on regional mean thickness estimates. All of this was made explicit in the introduction and method sections of our paper.

8 Review 3

This paper describes a new method for assessing cortical thickness from MR images. It is compared against an existing technique using a different approach using a set of normal images. The results are also compared against those from other studies in the literature measuring cortical thickness. The method described is shown to have some advantages compared to existing approaches. The method is applied to study the variation of cortical thickness with age.

Methods of measuring cortical thickness are valuable both in understanding brain anatomy and potentially in the diagnosis of diseases of the brain such as Alzheimer's disease. The paper is clearly written.

Specific comments

Introduction. If space allowed, perhaps by reducing the reference list, a typical image of a slice through the brain with the corresponding grey matter segmentation would be quite instructive.

Those images are available in the previous MIUA paper (2005) on the algorithm presented here. Repeating them would be a waste of space. The reference list is necessarily long: this paper describes a meta-study, and so incorporates results from a large number of other papers.

Results - first paragraph What are the errors in the mean differences on width between the regions e.g. Are they +/- 1SD ?

Is this not the usual way to quote errors?

Why should the random errors be related to the number of subjects?

We recommend that the reviewer familiarise themselves with any basic text on statistics: I can recommend R. Barlow, Statistics: A Guide to the use of Statistical Methods in the Physical Sciences, Wiley, 1989.

Figure 2a

it's really only a personal preference but I like to see graphs like this plotted with the y-axis starting at zero. They give a better impression of the relative change in the y value with variation in x i.e. in this case with age.

We disagree: producing graphs with a large amount of white space at the bottom is generally discouraged, as it makes the individual points harder to distinguish, particularly in a conference format such as the MIUA proceedings.

This is a somewhat surprising graph. Nearly all graphs I have seen of anatomical or physiological variables with age indicate a slow rate of change between 20 and 50 years of age with an increasingly rapid deterioration above 50 (eg. Nagata et al Neuroradiology 1987 29 327-332)

Whilst this is true of, for example, grey matter, white matter and total parenchymal volumes, it is not true of cortical thickness measurements. See, for example, Sowell et al. Nature Neuroscience (2003), Tompson et al., Neuroimage (2004), Toga et al., Trends in Neurosciences (in press). More worrying is the implied attitude underlying this comment. The results presented in the paper summarise measurements from 635 subjects, performed across multiple research groups, and using several different measurement techniques. Whilst the results may vary in detail, they all demonstrate that the rate of cortical thickness change reduces with age during adulthood. The reviewer appears to be criticising the results based on their subjective expectation, in the face of all available objective evidence.