

Quantifying cortical atrophy

ERIC TURKHEIMER, C MUNRO CULLUM, DONN W HUBLER, SYDNEY W PAVER,
RONALD A YEO, ERIN D BIGLER

From the University of Texas at Austin, Texas, USA

SUMMARY Most of the methods of quantifying cortical atrophy that have been proposed involve the estimation of the volume of enlarged sulci in the cerebral cortex. The authors propose that the surface area of the sulci is a more valid measure of cortical atrophy, and describe a system for measuring the surface area of the cortex, and present data in support of the method's reliability and validity.

Computed tomography (CT) makes it possible to assess brain atrophy quantitatively.¹ Efforts at quantification have focused in two areas: ventricular enlargement, and widening of the sulci and fissures, which is often referred to as "cortical atrophy." In general, methods for measuring ventricular size have proved more successful than those for measuring cortical atrophy. Despite problems arising from the partial volume artifact,² and slow progress from linear to planimetric to volumetric measurement³ most researchers have been satisfied with the validity of the several available methods for measuring ventricular volume.

Measurement of cortical atrophy has proved to be considerably more difficult. Jacob and Levy,² in fact, have rejected such measurements as too unreliable, and rely instead on clinical ratings of CT images. Other investigators have been disappointed with the absence of expected relationships between quantitative measurements of cortical atrophy and variables such as age and neuropsychological test performance.⁴⁻⁶ This paper proposes using the surface area of the brain as a measure of brain atrophy. Before proceeding to data concerning the reliability and validity of the proposed technique, other methods of quantifying atrophy that have been utilised will be reviewed briefly.

We will refer repeatedly in this paper to the validity of our and others' methods of quantification. In previous reports, several issues have sometimes been confounded in this area. One involves the distinction between the validity of a technique as a

measure of the physical atrophy that is visible in the scan image and the validity of cortical atrophy, however measured, as a predictor of the behavioral concomitants of dementia or head trauma. The former depends on the quality of the measurement technique, and is the usage that we will employ; the latter involves the empirical relationship between brain morphology and behaviour, and is less the concern of this paper.

Another area of confusion concerns the difference between the statistical validity of a measure, which may be directly assessed by the magnitude of the correlation between the measure and a criterion, and the diagnostic utility of a measure, which is more difficult to assess. Several authors⁷ have pointed out that some clearly demented patients present without any apparent cortical atrophy on CT, while other patients with severe atrophy do not present significant behavioural deficits. While this issue is important diagnostically, it does not limit the usefulness of quantitative CT measures in assessing the relationship between the degree of atrophy and the degree of impairment.

Three basic methods have been proposed for quantifying cortical atrophy. The first comprises various rating systems. These usually involve neurologists' or neuroradiologists' who rate the degree of atrophy from nonexistent to severe on a three, four or five point scale.⁷⁻¹² These rating systems have been among those that have shown weak relationships with behavioural variables^{10,11} although De Leon *et al*⁸ report more positive results. The relationship between ratings of atrophy and degree of impairment represents the second type of validity that was discussed above; one of the shortcomings of ratings of this type is that it is difficult to assess their validity in the other sense, that is, the extent to which the raters made accurate use of the information that was

Quantifying cortical atrophy

available in the scans. We will present some data relevant to this question below.

The reliability of ratings is also relevant, of course, but is rarely reported. De Leon¹¹ reports an inter-rater reliability of 0.89, which may account for their positive findings. This reliability may also be atypically high, because the development of the rating scale was a central purpose of their investigation.

The second, and most widely used means of measuring cortical atrophy is to measure the width of the four largest sulci and sum them.¹³⁻¹⁷ As is the case with ratings of atrophy, this method has shown mixed results in regard to relationships with dementia. Some of the same reasons may be cited. First, it is clear that selecting only the four largest sulci ignores a good deal of information present in the scan. The technique implicitly considers a scan with a few very large sulci as more atrophied than a scan with many moderately enlarged sulci, because sulci smaller than the fourth largest are ignored entirely. The reliability of this technique often is not reported, although Brinkman *et al* report an inter-rater reliability of 0.91.

The third means of measuring cortical atrophy has been developed by Jernigan and her co-workers.¹⁸⁻²⁰ This technique involves computing the number of low density pixels in the lateral cortex directly, using the digital version of the image that is generated by the CT scanner. Since each low density pixel approximates a voxel ($1 \times 1 \times 10$ mm in volume), the total low density volume can be computed from the total number of low density pixels. Jernigan *et al* use the total volume of cerebrospinal fluid near the surface of the brain as a measure of cortical atrophy. This approach may not be feasible for many researchers, however, because it requires the digital data from which the familiar film images are produced. In most settings only the serial CT images are retained as permanent data.

The Jernigan *et al* technique has produced high correlations between measured ventricle volume and neurologists' ratings of ventricular enlargement, but lower correlations between sulcal volume and

neurologists' ratings of cortical atrophy.¹⁸ Jernigan *et al* suggest that this low correlation may be due to bone artifact in areas of the scan image near the skull, or to the neurologists' unequal weighting of atrophy in different parts of the brain in making their ratings. We may add to these possibilities a third, that the construct of "atrophy" which the neurologists or neuroradiologists rate is inadequately measured by the volume of cerebrospinal fluid in the cortex.

When one examines a CT image for evidence of cortical atrophy, the most relevant feature is the degree of "convolutedness" of the surface of the brain. The widened sulci and fissures that produce these convolutions, however, often do not subtract significantly from the total volume of intact brain matter. Rather, convolutions will always have the effect of increasing the surface area of the brain, or, on a single slice, its circumference. We propose that measuring the surface area of the brain is the most valid way of quantifying the degree of atrophy that is visible in a CT scan.

The relationship between surface area and atrophy is illustrated in the figure, which is a computer generated representation of the brain and cranium from one slice of a highly atrophied brain. The total area of the sulci in this slice amounts to 18% of the total cranial area. The circumference of the cortex in the figure, however, is 253% greater than the circumference of the cranium. Thus, it appears that brain circumference, and hence surface area, may be a more sensitive measure of cortical atrophy than sulcal volume. We describe a method of making this measurement below.

Method

Subjects

Thirty subjects, including 10 patients with Alzheimer's disease, 10 patients with closed head injury, and 10 control subjects, were selected for analysis. The normal group consisted of five subjects similar in age to the trauma patients, and five subjects similar in age to the Alzheimer patients. The control subjects were within normal limits on

Table 1 Means and standard deviations of CT measurements

	Young		Old	
	Trauma (n = 10)	Controls (n = 5)	Alzheimer (n = 10)	Controls (n = 5)
Age	31.7 (11.9)	23.4 (8.4)	65.3 (18.4)	60.4 (7.4)
Atrophy index	7.4 (0.61)	7.0 (0.40)	8.1 (0.78)	7.0 (0.30)†
Widest sulci	2.2 (2.1)	0.8 (1.79)	3.7 (2.3)	3.2 (2.2)
VBR	0.06 (0.04)	0.02 (0.01)†	0.05 (0.03)	0.02 (0.01)†
Mettemyric fissure width	1.9 (1.1)	0.4 (0.89)*	1.7 (0.6)	1.4 (0.55)
Sylvian fissure width	1.8 (0.8)	0.8 (0.84)	2.0 (1.0)	2.0 (1.9)

Asterisks indicate the significance of t test comparisons between each patient group and its respective control group. *p < 0.05, †p < 0.01

Address for reprint requests: Dr E Turkheimer, Dept of Psychology, University of Texas at Austin, Austin, Texas 78712, USA.

Received 26 August 1983 and in revised form 13 April 1984.
Accepted 24 April 1984.

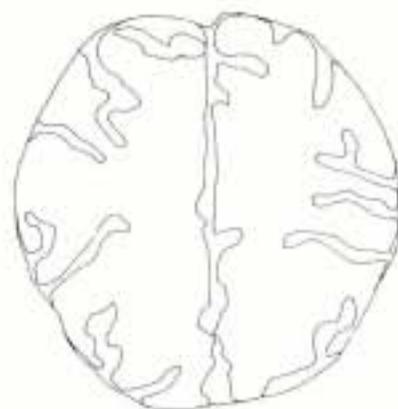


Fig. Computer drawn representation of slice from CT scan of a brain with significant cortical atrophy.

neurological and neuropsychological evaluation and had normal CT scans. The Alzheimer patients were diagnosed by board certified neurologists on the basis of neurological and neuropsychological examination. Mean ages of the subjects are given in table 1.

CT Scans

All scans were produced on an EMI 1010 scanner, with an 80×80 grid. The scans were placed on x-ray view boxes and the inner surface of the cranium, the perimeter of the brain, the ventricles and any subcortical low density areas (other than vascular and cisternal areas) were traced onto tracing paper by each of three of the authors, working independently. Eight consecutive slices were traced from each scan, starting with the first slice on which both temporal lobe tips were visible. In tracing the perimeter of the brain, special attention was given to tracing the details of all sulci that were clearly visible (see fig). The tracings were then digitised using a Summagraphics Bitpad by moving a cursor manually around the perimeter of all relevant structures. The Bitpad recorded the (x, y) coordinates of the cursor every 0.2 seconds.

The volumes and surface areas of the relevant structures were then computed from the digitised images using a program written in APLSF on a DEC-20 computer.²¹ The program computes the area and circumference of each structure in each slice of the CT series, and then computes volumes and surface areas by interpolating between slices using the trapezoidal rule.

The ventricle-brain ratio (VBR) was computed by dividing ventricle volume by cranial volume. Cranial volume was used instead of brain volume as it represents a more reliable estimate of head size in the presence of significant tissue loss in the brain. Because we did not include damaged tissue (for example, widened sulci and fissures) in our measure of brain volume, the volumes of atrophic brains were reduced relative to ventricle volume, independent of head size.

The surface area of the brain is not in and of itself an adequate measure of brain atrophy, because larger brains have more surface area. Therefore, the surface areas were

standardised by dividing them by the square root of the total cranial volume. The square root of the volume was used because the area of a structure in a single CT slice is a quadratic function of its circumference. The formula used to measure cortical atrophy, therefore, was:

$$\frac{\text{Brain Surface Area}}{\sqrt{\text{Cranial Volume}}}$$

We will refer to this value as the "atrophy index." It may be noted that since the volume of the brain is a cubic function of brain size while surface area is quadratic, the correct standardisation would appear to be the square root of surface area divided by the cube root of volume. However, because the number of slices used in computing the volume was held constant, and because the distance between slices is always the same, there is no variability in the third dimension, so this additional correction is not required.

The CT film images of the 20 patients were rated by two board certified neurologists on two five point scales, one for cortical atrophy and one for ventricular enlargement. Additionally, one of the authors and a colleague independently measured the widths of the four largest sulci, the greatest width of the interhemispheric fissure, and the greatest width of the sylvian fissure using a transparent ruler. A single rater made the same measurements on the control subjects.

Results

Means and standard deviations of the CT measurements are given in table 1. The significance levels presented refer to the *t* test comparisons between the trauma patients and the younger controls, and the Alzheimer patients and the older control subjects. The trauma patients differed significantly from the young controls only in terms of the VBR and interhemispheric fissure width measure. The atrophy index and other measures did not significantly differentiate these groups. Significant differences were observed on the Atrophy Index and the VBR. Other indices of cortical atrophy did not show significant differences between these two groups.

Reliability

Only the head trauma and the Alzheimer patients ($n = 20$) were used in the reliability analysis, the reliability of the measures was assessed using intraclass correlations (ICC), as described by Shrout and Fleiss.²² Two of their versions of ICC were employed. ICC-3 is comparable to the inter-rater correlations that usually are employed. Neither inter-rater correlations nor ICC-3 consider differences between the means of the raters' ratings as error. High correlations between the raters, therefore, produce high reliability coefficients regardless of consistent differences between the raters' mean rat-

Quantifying cortical atrophy

ings. ICC-2 is a more conservative index, in that mean differences are regarded as error. ICC-3 will be higher than ICC-2 to the extent that the raters differ as to the overall level of their ratings. All reported reliabilities represent the reliability of a single rater; higher reliabilities could be obtained by using the mean of several raters.

Table 2 presents the two reliability coefficients for each of the seven measures. Our measure of VBR has a reliability that approaches unity, as has been the case in previous reports.¹ Neurologists' ratings of ventricular atrophy are quite reliable also, though somewhat less so. The atrophy index has a reliability of 0.88 when assessed by ICC-3, and of 0.77 when assessed by ICC-2. As described above, this reflects the fact that while pairs of tracers were highly correlated, there were consistent differences between their means. The linear measures of sulcal and fissure width are much less reliable, but there is less error introduced by mean differences. The reliability of neurologists' ratings of cortical enlargement also is quite low. One may conclude from this that the atrophy index has superior reliability, but that one should have the same tracer prepare all scans for use within a single study, to avoid the introduction of error stemming from mean differences. It should also be noted that the tracers in this study worked completely independently of each other, without

Table 2 Reliability of cortical atrophy measures (method of measurement in parentheses)

	Intraclass correlation -2*	Intraclass correlation -3†
Atrophy index (computer)	0.77	0.88
VBR (computer)	0.99	0.99
Cortical rating (neurologists)	0.28	0.41
Ventricle rating (neurologists)	0.84	0.90
4 widest sulci (ruler)	0.68	0.69
Interhemispheric width (ruler)	0.66	0.66
Sylvian width (ruler)	0.33	0.33

*Mean differences between raters considered as error.
†Mean differences between raters not considered as error.

Table 3 Correlations among atrophy measures (method of measurement in parentheses) ($n = 30$)

	1	2	3	4	5	6
1 Atrophy index (computer)						
2 VBR (computer)	0.33					
3 Cortical rating (neurologists)	0.82	0.49				
4 Ventricle rating (neurologists)	0.35	0.93	0.55			
5 4 widest sulci (ruler)	0.69	0.34	0.68	0.42		
6 Interhemispheric width (ruler)	0.46	0.42	0.71	0.47	0.51	
7 Sylvian fissure width (ruler)	0.49	0.33	0.56	0.35	0.69	0.52

prior consensus about tracing conventions. Reaching agreement about tracing procedures beforehand might increase reliability substantially.

Relationships Among the Measures

Table 3 presents the correlations among the seven atrophy measures. The mean of all participants' ratings of each variable was used in this analysis, so the reliability of the measures is slightly higher than those presented above. The atrophy index has the highest correlation ($r = 0.82$) with neurologists' ratings of cortical atrophy, although it is not significantly higher than that between the four widest sulci and neurologists' ratings ($r = 0.68$). The atrophy index is relatively uncorrelated with VBR ($r = 0.33$).

The magnitude of the correlation between the atrophy index and sulcal width ($r = 0.69$) is surprising, in light of the relatively low reliability of the sulcal width measurement. It would appear that the width of the four widest sulci is a fair index of overall cortical atrophy, limited primarily by the low reliability introduced by the manual method of measurement. An automated technique of measuring sulcal width could therefore be expected to be a better measure of cortical atrophy.

Discussion

In establishing the usefulness of a new measure, the three critical issues are reliability, convergent validity and discriminant validity. The atrophy index is very reliable, due largely to the fact that it is highly automated. It demonstrates convergent validity in its high correlation with neurologists' clinical impressions of cortical atrophy, and discriminant validity in its relatively low correlations with ventricle volume, which appears to be a somewhat distinct aspect of brain atrophy.

The atrophy index also showed the expected mean differences between demented and normal older subjects, and was in agreement with neurologists' ratings in this regard. There was no significant difference between head trauma patients and a group of younger controls, while the neurologists' ratings did show a difference. Examination of the CT scans revealed that the atrophy in the Alzheimer group was characterised by widened sulci in all parts of the cortex; in the trauma group the cortical atrophy was predominant in the frontal regions, and especially in the interhemispheric fissure. This interpretation is supported by table 1, which shows that while the trauma patients had wider interhemispheric fissures than the Alzheimer patients, the Alzheimer patients had a higher mean atrophy index and wider sulci. This suggests that the

cortical atrophy that is often seen in Alzheimer's disease and head trauma may be qualitatively different to some degree.

The atrophy index requires fairly simple computer equipment and the CT film images. When these are available, it would appear to be superior to manual measurement of the sulcal widths. When such equipment is not available, sulcal widths may provide an acceptable alternative, especially if the mean of two independent measurements is used instead of a single measurement, in order to increase the reliability.

As discussed earlier, the validity of these techniques has been assessed in terms of their accuracy of measurement of the physical atrophy visible in the scan images, as opposed to their correlation with behavioural variables. It remains to be seen, of course, whether improvement in measurement techniques will lead to parallel improvements in the magnitude of observed brain-behaviour relationships. We intend to report data on this topic in subsequent papers.

The authors thank Douglas Hudson, MD and Michael Douglas, MD for rating the CT scans.

References

- Bird JM. Computerized tomography and dementia: a review. *Prog Neurobiol* 1982;19:91-115.
- Jacoby RJ, Levy R, Dawson JM. Computed Tomography in the Elderly: 1. The normal population. *Br J Psychiatry* 1980;136:249-55.
- Yeo RA, Turkheimer E, Bigler ED. Computer analysis of lesion volume: Reliability and utility. *Clin Neuropsychol* 1983;44:683.
- Fox JH, Kaszniak AW, Huckman M. Computerized tomographic scanning not very helpful in dementia—not in craniopharyngioma. *New Engl J Med* 1979;300:437.
- Ramani SV, Loewenson RB, Gold L. Computerized tomographic scanning and the diagnosis of dementia. *New Engl J Med* 1979;300:1336-7.
- Wells CE, Duncan GW. Danger of overreliance on computerized cranial tomography. *Am J Psychiatry* 1977;134:811-3.
- Jacoby RJ, Levy R. Computed Tomography in the Elderly 2. Senile Dementia: Diagnosis and Functional Impairment. *Br J Psychiatry* 1980;136:256-69.
- Roberts MA, Caird FL, Grossart KW, Steven JL. Computerized tomography and the diagnosis of cerebral atrophy. *J Neurol Neurosurg Psychiatry* 1976;39:909-15.
- De Leon MJ, George AE, Ferris SH, Reisberg B, Kircheff II. Computer tomographic evaluation of brain-behavior relationships in senile dementia. *Int J Neuroscience* 1981;12:246-7.
- Huckman MS, Fox J, Topel J. The validity of criteria for the evaluation of cerebral atrophy by computed tomography. *Neuroradiology* 1975;116:85-92.
- Ford CV, Winter J. Computerized axial tomography and dementia in elderly patients. *J Gerontol* 1980;36:164-9.
- Rieder RO, Mann LS, Weinberger DR, van Kammen DP, Post RM. Computed tomographic scans in patients with schizophrenia, schizoaffective, and bipolar affective disorder. *Arch Gen Psychiatry* 1983;40:735-9.
- De Leon MJ. Computed tomography evaluations of brain behavior relationships in senile dementia of the Alzheimer's type. *Dissertation Abstracts International* 1980;41:1308-9.
- Kohlmeyer K, Shamena AR. The size of the lateral ventricles and cortical sulci in old age subjects with and without dementia. *Neuroradiology* 1981;20:271-2.
- Brinkman SD, Sarwar M, Levin HS, Morris HH. Quantitative indexes of computed tomography in dementia and normal aging. *Neuroradiology* 1981;138:89-92.
- Gonzalez CF, Lantieri RL, Nathan RJ. The CT scan appearance of the brain in the normal elderly population: A correlative study. *Neuroradiology* 1978;16:120-2.
- Earnest MP, Heaton RK, Wilkinson WE, Manke WF. Cortical atrophy, ventricular enlargement and intellectual impairment in the aged. *Neurology (Minneapolis)* 1979;29:1138-43.
- Jernigan TL, Zatz LM, Naeser MA. Semiautomated methods for quantitating CSF volume on cranial computed tomography. *Radiology* 1979;132:463-6.
- Jernigan TL, Zatz LM, Feinberg I, Fein G. Measurement of cerebral atrophy in the aged by computed tomography. In: Poon LW, ed. *Aging in the 1980's: Psychological Issues*. Washington: American Psychological Association, 1980.
- Naeser MA, Gebhardt C, Levine HL. Decreased computerized tomography numbers in patients with pre-senile dementia: Detection in patients with otherwise normal scans. *Arch Neurol* 1980;37:401-9.
- Turkheimer E, Yeo RA, Bigler ED. Digital planimetry in APLSF. *Behavior Research Methods and Instrumentation* 1983;15:471-3.
- Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull* 1979;86:420-8.