Quantifying cortical atrophy
ERIC TURKHEIMER, C MUNRO CULLUM, DONN W HUBLER, SYDNEY W PAYER, 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. 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 and slow progress from linear to planimetric to volumetric measurements, 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, 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. 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 consequences 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 a 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 authors have pointed out that some clearly demarcated 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. These rating systems have been among those that have shown weak relationships with behavioural variables, although De Leon et al report some 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 raters made accurate use of the information that was 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 Leon et al report an inter-rater reliability of 0.89, which may account for their positive findings. This reliability may also be artificially 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. As is the case with ratings of atrophy, this method has shown mixed results in regard to relationships with dementia. Other examples of this problem 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 very large sulci as more atrophied than a scan with many moderately enlarged sulci, because sulci smaller that 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 that has been developed by Jernigan and her coworkers. 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 x 1 x 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 in the surface of the brain as a measure of 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. 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 convolution, and hence the “surface area” of the brain, 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 matched individually to the trauma patients, and five subjects similar in age to the Alzheimer patients. The control subjects were within normal limits on

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Means and standard deviations of CT measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Young (n = 10)</td>
</tr>
<tr>
<td>Age</td>
<td>(in years)</td>
</tr>
<tr>
<td>Trauma</td>
<td>M ± SD</td>
</tr>
<tr>
<td>Controls</td>
<td>M ± SD</td>
</tr>
<tr>
<td>Alzheimer</td>
<td>M ± SD</td>
</tr>
<tr>
<td>Controls</td>
<td>M ± SD</td>
</tr>
<tr>
<td>Atrophy Index</td>
<td>12 (17)</td>
</tr>
<tr>
<td>Cerebral ventricle volume</td>
<td>87 (79)</td>
</tr>
<tr>
<td>Sulcal area</td>
<td>12 (16)</td>
</tr>
<tr>
<td>Sulcal area index</td>
<td>0.85 (0.04)</td>
</tr>
</tbody>
</table>

Notes: Indicate the significance of t test comparisons between each patient group and its respective control groups.

- For 0.05, p < 0.01
adequate measure of brain atrophy, because larger brains
head size. were reduced relative to ventricle volume, independent of
have more surface area. Therefore, the surface areas were

measure of brain volume, the volumes of atrophic brains
aged tissue (for example, widened sulci and fissures) in our
were then computed from the digitised images using a
volumes and surface areas by interpolating between slices
using the trapezoidal rule.

The ventricle-brain ratio (VBR) was computed by divid-
ing ventricle volume by cranial volume. Cranial volume
was used instead of brain volume as it represents a more
accurate measure of the brain's total volume. The correct standardisation would appear to be the surface
area divided by volume of the brain. 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
corticocortical atrophy and one for ventricular enlarge-
ment. Additionally, one of the authors and a colleague inde-
dently 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 mea-
urements are given in table 1. The significance level
presented refer to the t-test comparisons between
the trauma patients and the younger controls, and the
Alzheimer patients and the older control sub-
jecteds. Trauma patients differed significantly from
the young controls and from the VBR of the interhemispheric fissure width measurements. The
atrophy index and other measures did not significantly
differentiate these groups. Significant differences
between the dementia group and older control
patients 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
were used in the reliability analysis. The relia-
blity of the measures was assessed using intra-
correlations (ICC), as described by Shrout and
Fleiss. Two of their versions of ICC were em-
ployed. ICC-3 is comparable to the inter-rater cor-
relations 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 in-
put concensus 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' rat-
ing of each variable was used in this analysis, so the
reliability of the measures is slightly higher than
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 unit 1. As has been
seen in previous reports, neurologists' ratings
of ventricular atrophy are quite reliable, also, though
generally less so. The index of cortical atrophy has a
reliability of 0.88, which is not as high as in previous
reports. The index of sulcal width has a reliability of
0.77, which is not as high as in previous reports.

The magnitude of the correlation between the
two measures is surprising, in that it is the highest
reliability introduced by the manual method of meas-
urement. An automated technique of measuring
cortical sulci would therefore be expected to be a
better measure of cortical atrophy.

Discussion
In establishing the usefulness of a new measure, the
critical task is to provide evidence of reliability, conver-
gent 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 impres-
sions of cortical atrophy, and discriminant validity
in its relatively low correlations with sulcal volume.
It appears that the wide sulci are a more conservative
index of overall cortical atrophy, limited primarily by the low rela-
From the results of this study, it appears that the
atrophy index is a more reliable measure of cortical
atrophy, because it is highly correlated with clinical
impressions of cortical atrophy, and because it is not
highly correlated with sulcal width, which is a more
conservative measure of cortical atrophy.

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

<table>
<thead>
<tr>
<th>Measure</th>
<th>Intrasubject correlation</th>
<th>Inter-subject correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atrophy index (computer)</td>
<td>0.77</td>
<td>0.88</td>
</tr>
<tr>
<td>VBR (computer)</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>Cortical rating (neurologists)</td>
<td>0.28</td>
<td>0.55</td>
</tr>
<tr>
<td>Ventricle rating (neurologists)</td>
<td>0.28</td>
<td>0.55</td>
</tr>
<tr>
<td>Sulcal width (VBR)</td>
<td>0.68</td>
<td>0.69</td>
</tr>
<tr>
<td>Sulcal width (VBR)</td>
<td>0.66</td>
<td>0.66</td>
</tr>
<tr>
<td>Sulcal width index</td>
<td>0.33</td>
<td>0.33</td>
</tr>
</tbody>
</table>

The atrophy index also showed the expected
differences between the normal and normal
older subjects, and was in agreement with neurolog-
ists' 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

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

Turkheimer, Munro Callum, Hubler, Payer, Yeo, Blyer.
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