Neuroradiology (1994) 36: 17–19

Neuro radiology

Interobserver variability in CT assessment of brain atrophy

M. Leonardi[1], S. Ferro[2], R. Agati[3], L. Fiorani[2], A. Righini[3], E. Cristina[2], R. D'Alessandro[2]

[1] Neuroradiological Department, Udine Hospital, Udine, Italy
[2] Institute of Neurology, Neuroepidemiology Unit, University of Bologna, Bologna, Italy
[3] Neuroradiological Unit, University of Bologna, Bologna, Italy

Received: 20 October 1992/Accepted: 23 February 1993

Abstract. To assess interobserver variability in estimation of brain atrophy based on CT, four neuroradiologists examined CT brain images of 150 consecutive patients without focal lesions. An independent neuroradiologist made the following quantitative measurements: frontal horn index, subarachnoid space area and the ratio between subarachnoid space area and inner skull space area. Level of agreement was fair for the presence (k = 0.24), slight for the degree (mild, moderate, severe) (k = 0.24) and moderate for the type (cortical, subcortical, mixed) of atrophy (k = 0.59). There was a highly significant correlation between the number of observers agreeing and quantitative measurements. We concluded that neuroradiologists' subjective estimation of brain atrophy alone is not reliable. Quantitative measurements would be needed in cases where the presence of brain atrophy might determine clinical decisions.

Key words: Brain atrophy – Interobserver variability – Computed tomography

"Brain atrophy" is a term commonly used in clinical practice to indicate an apparent of loss of brain volume on CT. It is commonly described as consisting of enlargement of sulci, fissures and ventricles [1–3]. Such pictures may be seen both in pathological states and in healthy individuals [4–6].

Many studies have attempted to establish quantitative measures of pathological brain atrophy, with conflicting results [7–15], so that identification of atrophy on CT has been based on the experience of the neuroradiologist, at least in clinical practice.

Although good agreement has been shown for ventricular dilatation [16], no studies have evaluated interobserver variability in the estimation of cerebral atrophy. We investigated this variability and correlated it with quantitative data.

Correspondence to: M. Leonardi, Servizio di Neuroradiologia, Ospedale Civile di Udine, I-33100 Udine, Italy

Methods

The study was performed on 150 consecutive outpatients during 1990 (79 men, mean age 62.2, range 40–89; 71 women, mean age 62.9, range 43–93).

All patients with focal lesions visible on CT were excluded. CT sections were obtained in an axial plane at approximately 20° to the canthomeatal line, using a 320 × 320 matrix, 120 kVp and 320 mA. Section thickness was 5 mm with a 10 mm gap.

The images were interpreted separately, without any quantitative measurements, by four expert neuroradiologists from the same team, unaware of the patient's condition. They answered the following question: atrophy yes/no? If yes: mild, moderate, severe and cortical, subcortical or mixed?

Quantitative data were subsequently measured for each patient by an independent neuroradiologist:

1. Frontal horn diameter (FHD): the distance between the tips of the frontal horns measured on the cut showing them best [4,7–10,13,15–18].
2. Inner table diameter (ITD): the distance between the inner tables of the skull measured on the same cut, along the same line as the FHD [8, 9, 11, 13, 15–17].
3. Frontal horn index (FHI): the ratio of FHD to ITD [8,9,11,13,15–18].
4. Inner skull space area (ISSA): the area enclosed by the inner table of the skull [9, 12, 18, 19] calculated by outlining the contour of the inner table of the skull on the first cut above the bodies of the lateral ventricles.
5. Brain area (BA): the area of the brain tissue [9,12,18,19] calculated by outlining the contour of the brain tissue on the same cut as the ISSA.
6. Subarachnoid space area (SSA): obtained by subtraction of the BA from the ISSA [9, 12, 18, 19].
7. Ratio of SSA to ISSA [20].[1]

Table 1. k ranges and levels of agreement

k	Strength of agreement
< 0.00	Poor
0.00–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost perfect

[1] The ratio between the subarachnoid space area and the inner skull space area (SSA/ISSA) can be considered as having the same meaning of the complementary, opposite, ratio between brain area and inner skull space area: SSA + BA = ISSA.

Table 2. Distibution of subjective assessments of cerebral atrophy. (k values indicate general interobserver agreement)

	Atrophy present		Mild		Moderate		Severe		Subcortical		Cortical		Mixed	
	No	(%)	No	(%)	No	(%)	No	(%)	No	(%)	No	(%)	No	(%)
Observer 1	89	(59.4)	49	(55.1)	31	(34.8)	9	(10.1)	6	(6.7)	65	(73.0)	18	(20.2)
Observer 2	126	(84.0)	57	(45.2)	57	(45.2)	12	(9.5)	4	(3.1)	87	(69.0)	35	(27.7)
Observer 3	36	(24.0)	19	(52.7)	11	(30.5)	6	(16.7)	8	(22.3)	18	(50.0)	10	(27.8)
Observer 4	50	(33.4)	26	(52.0)	20	(40.0)	4	(8.0)	6	(12.0)	30	(60.0)	14	(28.0)
k values		0.24			0.20[a]						0.59[a]			

[a] k values computed on 29 cases considered as showing atrophy by all observers

K statistics were used as a measure of interobserver agreement. They take into account and correct for chance-explained agreement [21, 23]. The formula applied is

$$k = (Po - Pe)/(1 - Pe)$$

where Po is the crude proportion of agreement and Pe is the chance-explained proportion of agreement. Estimates with their standard errors (SE) were computed according to Fleiss [21], providing a measure of agreement among multiple raters. Statistical significance of k value may be tested, considering that the quantity Z = k/SE (k) (k value divided by its SE) has an asymptotically Gaussian distribution. For complete agreement, k = 1.00; for a completely chance-explained agreement, k = 0.00. A negative k value significantly different from zero suggests systematic disagreement [22].

The interpretation of the k values is shown in Table 1.

K statistics were computed for general agreement, for each pair of observers and for each category (degree and location of atrophy) only in cases considered to show brain atrophy by all observers.

To establish whether the judgment as to the presence of atrophy by the neuroradiologists could be associated with quantitative measurements, we considered only the FHI, SSA and the SSA/ISSA ratio. We also took into account the terzili of the frequency distribution of these quantitative variables according to the numbers of observers judging atrophy to be present (Noss variable).

Chi-square test for trend was used as an association index between the Noss variable and quantitative data.

Table 3. Interobserver agreement and k values for presence of atrophy for different pairs of ($n = 150$)

Pairs of observers	Agreement %	k values
1–2	71	0.34
1–3	61	0.29
1–4	69	0.42
2–3	40	0.11
2–4	48	0.14
3–4	83	0.59

k overall = 0.34; chi square (df = 5) = 29.6; $p < 0.001$

Results

The assignment to each diagnostic category made by each neuroradiologist is shown in Table 2. The percentages of patients judged to have atrophy ranged from 24% to 84%. The level of agreement among radiologists for each of the specific categories is indicated by the K values. The chance-corrected level of agreement was fair for the presence of brain atrophy (k = 0.24), slight for its degree (k = 0.20) and moderate for its type (k = 0.59).

Separate analysis of interobserver agreement between different combinations of observers is shown in Table 3. The chance-corrected level of agreement ranged from a minimal value of k = 0.11 (slight) to a maximum of k = 0.59 (moderate).

Quantitative measurements (mean, SD, range) are shown according to the number of observers agreeing on atrophy (Table 4).

The associations between quantitative measurements and clinical estimation of atrophy was highly significant for FHI (chi square (df 1) = 21.14; $p < 0.0001$), SSA (chi square = 35.34; $p < 0.0001$) and the SSA/ISSA ratio (chi square = 33.55; $p < 0.0001$).

Table 4. Quantitative CT data and number of observers on presence of atrophy

Noss[a]	No cases	Frontal horn index	Subarachnoid space area	SSA/ISSA ratio[b]
		Mean SD (range)	Mean SD (range)	Mean SD (range)
0	20	30.6 3.02 (24.6–36.5)	2159.06 623.26 (1318.4–3577.6)	15.76 5.12 (10.3–25.6)
1	38	30.47 2.91 (24.8–37.3)	2693.70 794.15 (1760.6–4508.2)	18.59 5.7 (11.7–18.6)
2	43	32.59 3.61 (26.6–40)	3166.26 732.10 (1575.7–4831.4)	22.32 5.91 (10.2–35.4)
3	20	33.27 3.47 (28.1–41.8)	3230.26 670.93 (2194.6–4547.2)	24.28 4.57 (15.8–32.7)
4	29	36.87 6.46 (29–53.8)	3821.72 1025.17 (1775.4–6379)	28.07 7.72 (12.4–46.8)

[a] Noss: number of observers judging atrophy to be present
[b] SSA/ISSA: ratio of subarachnoid space area to inner skull space area

Discussion

To our knowledge no previous studies on interobserver agreement in clinical estimation of brain atrophy on CT have been performed. If this diagnosis might affect management it would be important to know the reliability and reproducibility of radiological assessment. We found that in general interobserver agreement was fair. Further analysis of agreement between pairs of observers was at best moderate. Comparison with quantitative measurements was performed with only the FHI, for which normal values are in the literature [2, 9, 15, 24]. Although there is a general increase in mean values with the number of observers agreeing on brain atrophy, analysis of standard deviation and range (Table 4) shows that some subjects judged as having no atrophy had in FHI of more than the normal value of 30 %; others, thought to show by three or four observers brain atrophy had a normal FHI. This did not apply to the other measures, for which we had no standardised value. However, the chi square trend analysis showed a high correlation between the number of observers agreeing as regards all measures considered.

Radiological estimation of atrophy is not reliable. Should the question of brain atrophy be relevant to management decisions, quantitative measurements should improve diagnostic reliability. We used the FHI and the SSA/ISSA ratio. The first measures lateral ventricle dilatation and correlates well with brain atrophy, whereas the second gives an index of enlargement of the subarachnoid space. Although standardised values are given for the FHI [15], no previous studies have standardised an index of subarachnoid space enlargement. We found a strong correlation between the SSA/ISSA ratio and agreement on subjective assessment of atrophy, particularly at SSA/ISSA levels of more than 25. Further studies would be useful to obtain standardised values for these measures.

References

1. Moseley I (1986) Diagnostic imaging in neurological disease. Churchill Livingstone, Edinburgh
2. LeMay M (1984) Radiologic changes of the aging brain and skull. AJNR 5: 269–275
3. Leonardi M, Martelli A, Costa A, Mauri M, Zanotti B, Zappoli F (1991) Imagerie cérébrale: applications neuropathologiques à la maladie d'Alzheimer: le rôle de la TDM et evaluation endocrine. Bull Assoc Anat 75: 97–99
4. Kohlmeyer K, Shamena AR (1983) CT assessment of CSF spaces in the brain in demented and nondemented patients over 60 years of age. AJNR 4: 706–707
5. Laffey PA, Peyster RG, Nathan R, Haskin ME, McGinley JA (1984) Computed tomography and aging: results in a normal elderly population. Neuroradiology 26: 273–278
6. LeMay M (1986) CT changes in dementing diseases: a review. AJNR 147: 963–975
7. Nagata K, Basugi N, Fukushima T, et al (1987) A quantitative study of physiological cerebral atrophy with aging. A statistical analysis of the normal range. Neuroradiology 29: 327–332
8. Gomori JM, Steiner I, Melamed E, Cooper G (1984) The assessment of changes in brain volume using combined linear measurements. A CT-scan study. Neuroradiology 26: 21–24
9. Sabattini L (1982) Evaluation and measurement of the normal ventricular and subarachnoid spaces by CT. Neuroradiology 23: 1–5
10. Hirashima Y, Shindo K, Endo S (1987) Measurement of the area of the anterior horn of the right lateral ventricle for the diagnosis of brain atrophy by CT. Neuroradiology 25: 23–27
11. Adam P, Fabre N, Guell A, Bessoles G, Roulleau J, Bès A (1983) Cortical atrophy in Parkinson disease: correlation between clinical and CT findings with special emphasis on prefrontal atrophy. AJNR 4: 442–445
12. Yerby MS, Sudsten JW, Larson EB, Wu SA, Sumi SM (1985) A new method of measuring brain atrophy: the effect of aging in its application for diagnosing dementia. Neurology 35: 1316–1320
13. Steiner I, Gomori JM, Melamed E (1985) Features of brain atrophy in Parkinson's disease. A CT scan study. Neuroradiology 27: 158–160
14. Lee D, Fox A, Viñuela F, et al (1987) Interobserver variation in computed tomography of the brain. Arch Neurol 44: 30–31
15. Soininen H, Puranen M, Riekkinen PJ (1982) Computed tomography findings in senile dementia and normal ageing. J Neurol Neurosurg Psychiatry 45: 50–54
16. LeMay M, Stafford JL, Sandor T, Albert M, Haykal H, Zamani A (1986) Statistical assessment of percentual CT scan ratings in patients with Alzheimer type dementia. J Comput Assist Tomogr 10: 802–809
17. Inzelberg R, Treves T, Reider I, Gerlenter I, Korczyn AD (1987) Computed tomography brain changes in Parkinsonian dementia. Neuroradiology 29: 535–539
18. De Leon MJ (1989) Alzheimer's Disease: longitudinal CT studies of ventricular change. AJNR 10: 371–376
19. Arai H, Kobayasci K, Ikeda K, Nagao Y, Ogiara R, Kosaka K (1983) A computed tomography study of Alzheimer's disease. J Neurol 229: 69–77
20. Agati R, D'Alessandro R, Fiorani L, Righini A, Leonardi M (1992) Valutazione quantitativa dell'atrofia cerebrale in tomografia computerizzata. Rivis Neuroradiol 5: 185–193
21. Fleiss JH (1971) Measuring nominal scale agreement among many raters. Psychol Bull 76: 378–382
22. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33: 159–174
23. Fleiss JL (1981) Statistical methods for rates and proportion. Wiley, Louton, pp 212–234
24. Von Gall M, Artmann H, Lerch G, Nemeth N (1978) Results of computed tomography on chronic alcoholics. Neuroradiology 16: 329–331