Evaluation of Selection Criteria Used in Alzheimer's Disease Clinical Trials

J. Willmer and E. Mohr

ABSTRACT: ***Background:*** In the absence of a biological marker for Alzheimer's disease (AD), diagnosis has to be achieved using clinical criteria sets such as those outlined in DSM-IV, NINCDS-ADRDA, or ICD-10. As these criteria are quite broadly defined, there may be inter-rater variability in interpretation. ***Methods:*** Using a previously published CT scan measuring technique which correlates well with diagnoses achieved using the NINCDS-ADRDA criteria as interpreted at our clinic, we chose to independently examine and reach a diagnosis in patients selected for participation in clinical trials of therapeutic agents for the treatment of AD. Forty-four CT scans from six investigators across Canada were examined using this model. All patients had been diagnosed as having AD by NINCDS-ADRDA criteria and were deemed acceptable to participate in a clinical trial. ***Results:*** The diagnostic concordance achieved in the original published model was 91.5%. The diagnostic concordance in the population currently being studied was 77.3%. However when examined by site, results ranged from 57.1% to 100%. Using the model, an index of atrophy and a probability of diagnosis of AD can be determined. Across sites, there were statistically significant differences in these measures ($p \leq 0.035$). The mean probability of diagnosis of AD across sites ranged from 0.56 to 0.94. Although the sites with lower probabilities had slightly lower mean ages and slightly less atrophy, there was no overall correlation of the atrophy measures with age. ***Conclusions:*** Current results raise the possibility that the selection of patients for AD clinical trials using current diagnostic criteria sets may not be adequate and conclusions with respect to agent efficacy could be flawed.

RÉSUMÉ: Évaluation des critères de sélection utilisés pour les essais cliniques dans la maladie d'Alzheimer. ***Introduction:*** Comme il n'y a pas de marqueur biologique de la maladie d'Alzheimer (MA), le diagnostic est posé au moyen de critères cliniques dont ceux du DSM-IV, du NINCDS-ADRDA, ou du ICD-10. Comme ces critères ont une dèfinition large, il peut y avoir une variabilité interobservateur dans leur interprétation. ***Méthodes:*** Nous avons examiné indépendamment des patients sélectionnés pour participer à des essais de pharmacologie clinique pour le traitement de la MA et nous avons posé un diagnostic en utilisant une technique déjà publiée de mesure par CT scan dont la corrélation avec le diagnostic posé au moyen des critères du NINCDS-ADRDA tels qu'interprétés à notre clinique est bonne. Quarante-quatre CT scans fournis par six investigateurs à travers le Canada ont été examinés au moyen de ce modèle. Tous les patients avaient reçu un diagnostic de MA, selon les critères du NINCDS-ADRDA et étaient considérés comme admissibles à un essai clinique. ***Résultats:*** La concordance diagnostique réalisée dans le modèle original publié était de 91.5%. La concordance diagnostique dans l'échantillon que nous avons étudié était de 77.3%. Cependant, quand les résultats étaient étudiés par centre, les résultats variaient de 57.1% à 100%. Un index d'atrophie et la probabilité du diagnostic de MA peuvent être déterminés au moyen de ce modèle. Il y avait des différences statistiquement significatives dans ces mesures ($p \leq 0.035$) entre les centres. La probabilité moyenne d'un diagnostic de MA variait de 0.56 à 0.94, selon les centres. Bien que les patients des centres qui avaient une probabilité diagnostique plus faible avaient un âge moyen légèrement plus bas et un peu moins d'atrophie, il n'y avait pas en général de corrélation des mesures d'atrophie avec l'age. ***Conclusions:*** Ces résultats soulèvent la possibilité que, pour les essais thérapeutiques, la sélection des patients au moyen des critères diagnostiques en usage courant ne soient pas adéquats et que les conclusioins quant à l'efficacité de ces agents soient inexactes.

Can. J. Neurol. Sci. 1998; 25: 39-43

Alzheimer's disease (AD) is becoming more prevalent as our population ages. This is due to the fact that people are living longer and that the population therefore has a larger proportion of older individuals. Advancing age is the greatest risk factor for the development of AD.[1-3] Although recent therapeutic advances hold considerable promise, [4-12] there is still no definitive therapy for this disease. There are many new medications on the horizon that are based on interventions in the known pathophysiology of AD. In order to properly evaluate these, an accurate diagnosis of individuals with this illness is imperative.

In the absence of biological markers for AD, diagnosis has to be achieved using clinical criteria sets such as those outlined in DSM-IV, NINCDS-ADRDA, or ICD-10.[13-16] These diagnostic check lists use broad criteria that could be open to interpretation by different clinicians, especially when one considers that these

From the Division of Neurology, University of Ottawa, SCOHS - EB Pavilion.
RECEIVED APRIL 9, 1997. ACCEPTED IN FINAL FORM AUGUST 11, 1997.
Reprint requests to: Jonathan Willmer, 75 Bruyere Street, Ottawa, Ontario, Canada K1N 5C8

individuals may have widely varying types and depth of training. In centres with a great deal of experience in the diagnosis of AD, the criteria are reliable,[17,18] but this may not be the case in centres with relatively less experience. In such instances, general application of these criteria may lead to misdiagnosis, particularly in disorders such as some of the frontal lobe dementias and Lewy body dementia.[19] This may introduce sufficient variability to distort patient selection for clinical trials, which may not be consistent across sites in a multicentre trial.

Although autopsy confirmation would be the most reliable test of these criteria sets, it is not always possible to obtain. Previous clinical neuropathological correlation studies have generally shown good correlation of the autopsy findings with the clinical criteria.[20-26] These studies were done in large specialty centres with clinical diagnosis often reached by a team consensus. Perhaps the best diagnostic-neuropathologic correlations are obtained in the centres involved in the CERAD protocols.[26] This is likely due to the standardization of clinical and neuropathological assessments. With the rapid development of new medications to be tested in dementias, new clinical sites are being recruited at an increasing rate. Some of these sites may not have as much diagnostic experience as those involved in the initial correlation studies.

There has been some discussion, particularly in the CERAD groups, of standardizing radiological protocols. Davis et al.[27] have shown good correlation of specific radiological measurements, particularly those of the temporal lobes, with neuropathological findings in AD. Jobst et al.[28-30] have shown good clinical and neuropathological correlation with a CT scan model that looks at temporal lobe structures.

Based on these contentions, we decided to test the NINCDS-ADRDA criteria in a number of clinical centres across Canada, against an independent tool, a previously published CT scan measuring technique.[31] This model correlates well with diagnoses achieved using the NINCDS-ADRDA criteria as interpreted at our site. A key measurement in this model is the diameter of the temporal horns. Davis et al.[27] have shown that temporal horn enlargement in particular is important in individuals with AD.

Methods

Using the CT scan model, we chose to independently examine and reach a diagnosis in patients selected for participation in clinical trials of therapeutic agents in AD. The CT slices and measurements used in the model are identical to those outlined by Willmer et al.[31] The diagnostic concordance achieved in the original published model was 91.5%.

Forty-four CT scans from six sites across Canada were examined using the model. All patients had been diagnosed as having probable AD by NINCDS-ADRDA criteria and were deemed acceptable to participate in a clinical trial. Age and sex distribution of the groups is shown in Table 1. There was no statistically significant difference in age across the groups.

A standard scanning protocol was used at all sites. Scans were conducted at an angle of positive 20 degrees to the canthomeatal line to get better visualization of the temporal lobes and the temporal horns.[32] A series of cuts five millimetres thick spaced at five millimetre intervals were obtained, extending from the cranial floor to the first cut which did not contain any portion of the lateral ventricles. Three slices were chosen for measurements. Slice 1 was chosen as that showing the widest

Table 1: Age and Sex by Site.

Site	Age ± Std. Dev. (yrs.)	Male/Female
A	71.3 ± 3.6	3/3
B	78.0 ± 6.5	2/4
C	67.0 ± 7.4	2/5
D	72.1 ± 6.0	6/2
E	72.5 ± 4.2	1/3
F	68.0 ± 8.1	4/9

cross section of the temporal horns, slice 2 showed the widest spacing of the frontal horns and slice 3 showed the greatest measurement of the lateral ventricular waist. Three measurements are necessary for the diagnostic equation. One investigator (JW) made all measurements. From slice one, the sums of the left and right temporal horns were measured perpendicularly to the longest axis seen on the CT scan at the widest point. From slice 2, the widest distance between the caudate nuclei at their midpoints was measured. From slice 3, the narrowest point or ventricular waist was measured. All measurements were expressed as ratios of the internal skull diameter (from inner table to inner table) measured horizontally at the same level as the measurement. This gave three variables named TH, BC, and LV respectively. From the original model, classification equations can be defined as follows:

$$A = (-6.6)\,TH + (235.19)\,BC + (88.94)\,LV - 41.91$$
$$N = (-81.54)\,TH + (208.98)\,BC + (51.46)\,LV - 24.99$$

A diagnosis of AD is predicted when A > N. Furthermore, the probability of diagnosis of AD can be calculated as follows:

$$P = exp A / (exp A + exp N)$$

An atrophy index was calculated based on the three ratios as follows:

$$AI = TH + BC + LV$$

Tests of difference were applied to the six sites for the variables; diagnostic concordance, probability of diagnosis of AD and atrophy index. Due to the small and unequal groups in the study, nonparametric tests of difference were chosen. The Kruskal-Wallis test was used to look at over all group differences, and the Kolmogorov-Smirnov test was used for pairwise comparisons.

Results

The diagnostic concordance between the NINCDS-ADRDA diagnosis and that obtained from the CT model was 77.3% overall. However when examined by site, results ranged from 57.1% to 100% (Table 2). The atrophy index and probability of diagnosis of AD are also shown. The mean probability of diagnosis of AD across sites ranged from 0.56 to 0.94. There were statistically significant differences in these measures ($p \leq 0.035$) using the Kruskal-Wallis test. The means for each site for the two variables are shown graphically in Figures 1 and 2.

The differences between pairs of sites for these two variables can be calculated using the Kolmogorov-Smirnov test. The

https://doi.org/10.1017/S0317167100033461 Published online by Cambridge University Press

Table 2: Diagnostic Accuracy, Mean Probability of Diagnosis of AD and Mean Atrophy Index by Site.

Site	Concordance (%)	Probability*	Atrophy Index**
A	100.0	0.94	0.58
B	83.3	0.91	0.60
C	57.1	0.56	0.45
D	100.0	0.93	0.58
E	75.0	0.78	0.54
F	61.5	0.63	0.50

* Differences significant at p = 0.035
** Differences significant at p = 0.026

probabilities of difference between pairs of sites are shown in Tables 3 and 4. Although the sites with lower probabilities had slightly lower mean ages and slightly less atrophy, there was no overall correlation of the atrophy measures with age.

Discussion

Current results suggest that there are substantial differences in Alzheimer patient populations between sites when the pattern of cerebral atrophy as measured by our CT scan model is used. A number of explanations for these discrepancies can be suggested. Some of the variability between sites could be explained by the fact that the scans were all done on different scanners. The CT scan model maybe measuring something different than the clinical criteria alone, leading to selection of different subpopulations. Thus, different subgroups of AD maybe present at different sites. The CT model reflects structural changes while the clinical criteria generally reflect functional changes. There is correlation between the two; however, some cases do not fit well. The changes seen on the CT scan reflect atrophy, which is

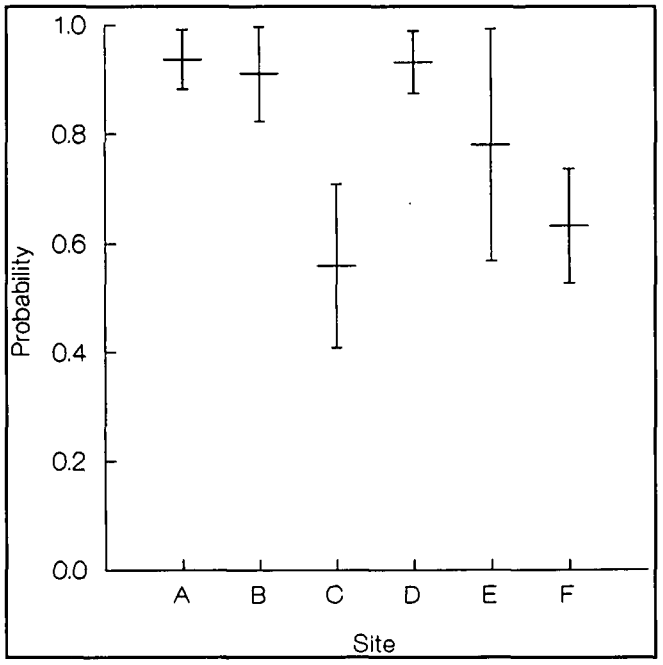

Figure 1: Mean probability of diagnosis of AD by site (showing standard error bars).

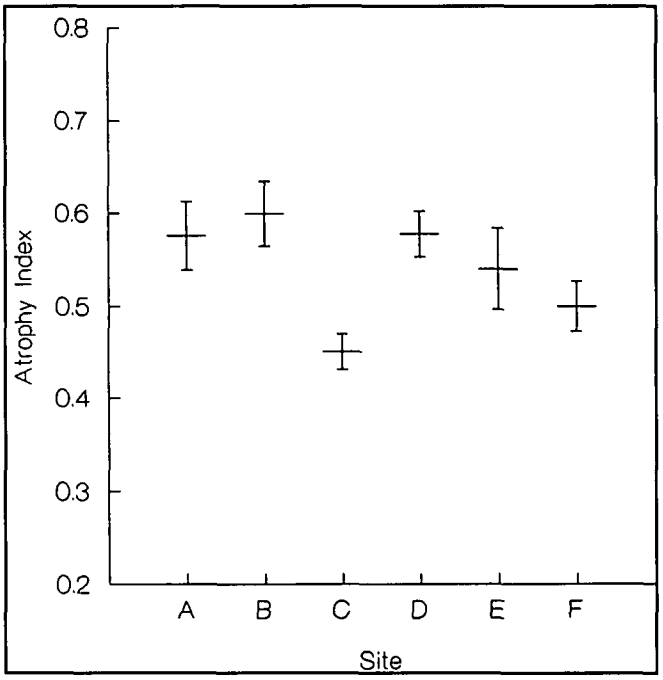

Figure 2: Mean atrophy index by site (showing standard error bars).

a neuropathological process. This highlights the need for eventual autopsy correlation of the model. At this point in time, only one of the original subjects in the initial study[31] has been followed until death. AD was neuropathologically confirmed. There is however support for this type of approach using CT scanning, and there is correlation with neuropathological changes in the temporal lobe.[27,28]

While neuropathological validation of the CT scanning model is important, the aim of the present investigation was to use this model as an independent assessment tool. Data supporting the notion that the model correlates well with our own interpretation of the NINCDS-ADRDA criteria have been published previously.[31] Hence any differences noted in the interpretation between groups would indicate that other sites participating in the current study were perhaps interpreting the diagnostic criteria differently than us.

The different sites may be applying the NINCDS-ADRDA criteria differently, again resulting in different subsets of AD, or in the worst case, perhaps not all with equal accuracy. This may result in some sites having a lower proportion of true AD patients. These points suggest possible diagnostic implications. The model may be a useful adjunct in the diagnosis of AD, and

Table 3: Probability of Similarity Between Pairs of Sites for Probability of Diagnosis of AD.

Site	A	B	C	D	E	F
A	1.000	–	–	–	–	–
B	0.333	1.000	–	–	–	–
C	0.009	0.009	1.000	–	–	–
D	1.000	0.778	0.004	1.000	–	–
E	0.944	0.944	0.125	0.875	1.000	–
F	0.149	0.149	0.265	0.098	0.479	1.000

https://doi.org/10.1017/S0317167100033461 Published online by Cambridge University Press

Table 4: Probability of Similarity Between Pairs of Sites for Atrophy Index.

Site	A	B	C	D	E	F
A	1.000	–	–	–	–	–
B	0.778	1.000	–	–	–	–
C	0.009	0.009	1.000	–	–	–
D	0.906	0.437	0.004	1.000	–	–
E	0.944	0.500	0.125	0.875	1.000	–
F	0.149	0.149	0.265	0.132	0.479	1.000

particularly, may be useful in patient selection for therapeutic trials in AD. Non-clinical criteria such as those in the CT model could be used to supplement the diagnostic criteria rather than replace them. In light of the previous discussion, using these criteria in addition to the clinical ones may better reflect the underlying neuropathology of AD. A follow up study is being planned to re-evaluate the diagnoses given to these patients at the start of the clinical trial by examining them again at least twelve months later. These data will then be compared to the diagnosis achieved by the model to see if there has been a change. Ultimately, however, the true test will be to follow these patients to autopsy.

It is still unknown how the model as a diagnostic tool compares with other biological markers. Until recently their sensitivity has been rather low but newer indicators such as apolipoprotein E (Apo-E) and P97 may be more selective.[33-35] It will be important to determine if the differences between our CT model and the criteria based diagnosis are reflected in one or more of these markers.

If these explanations are valid, the selection of patients for Alzheimer clinical trials using current diagnostic criteria sets may not be adequate and conclusions with respect to agent efficacy could be flawed. As the NINCDS-ADRDA criteria have been shown to be reliable in expert hands and that there may be correlation of the CT model with diagnosis achieved using them, it may well be a useful adjunct in accurately and consistently selecting patients for clinical trials. The model uses scanning parameters close to those used in standard clinical scans, and hence should be easily performed in any centre with a CT scanner. The measurement techniques can be easily learned and we believe them to be reliable and reproducible. We recommend that selection criteria for future clinical trials utilize this model as an adjunct to the clinical diagnosis of AD.

Acknowledgements

The authors thank all participating sites from the Consortium of Canadian Centers for Clinical Cognitive Research (H. Feldman - Vancouver, D. Hogan - Calgary, H. Masson - Montreal, and P. Bailey - Saint John).

References

1. Mohr E, Feldman H, Gauthier S. Canadian guidelines for the development of antidementia therapies: a conceptual summary. Can J Neurol Sci 1995; 22: 62-71.
2. Anonymous. The Canadian Study of Health and Aging: risk factors for Alzheimer's disease in Canada. Neurology 1994; 44: 2073-2080.
3. Ebly EM, Parhad IM, Hogan DB, Fung TS. Prevalence and types of dementia in the very old: results from the Canadian Study of Health and Aging. Neurology 1994; 44: 1593-1600.
4. Ehlert FJ, Roeske WR, Yamamura HI. Muscarinic receptors and novel strategies for the treatment of age-related brain disorders. Life Sci 1994; 55: 2135-2145.
5. Fisher A, Heldman E, Gurwitz D, et al. M1 agonists for the treatment of Alzheimer's disease. Novel properties and clinical update. Ann N Y Acad Sci 1996; 777: 189-196.
6. Hock FJ. Therapeutic approaches for memory impairments. Behav Brain Res 1995; 66: 143-150.
7. Madden S, Spaldin V, Park BK. Clinical pharmacokinetics of tacrine. Clin Pharmacokinet 1995; 28: 449-457.
8. Rother M, Kittner B, Rudolphi K, Rossner M, Labs KH. HWA 285 (propentofylline) – a new compound for the treatment of both vascular dementia and dementia of the Alzheimer type. Ann N Y Acad Sci 1996; 777: 404-409.
9. Whitehouse PJ, Voci J. Therapeutic trials in Alzheimer's disease. Curr Opin Neurol 1995; 8: 275-278.
10. Breitner JC. The role of anti-inflammatory drugs in the prevention and treatment of Alzheimer's disease. Annu Rev Med 1996; 47: 401-411.
11. Crismon ML. Tacrine: first drug approved for Alzheimer's disease. Ann Pharmacother 1994; 28: 744-751.
12. Delumeau JC, Bentue-Ferrer D, Gandon JM, et al. Monoamine oxidase inhibitors, cognitive functions and neurodegenerative diseases. J Neural Transm Suppl 1994; 41: 259-266.
13. McKhann MD, Drachman D, Folstein M, et al. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 1984; 34: 939-944.
14. World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines. World Health Organization: 1992.
15. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: Revised, 3rd Edition. Washington, D.C.: American Psychiatric Association; 1987.
16. Kukull WA, Larson EB, Reifler BV, et al. The validity of 3 clinical diagnostic criteria for Alzheimer's disease. Neurology 1990; 40: 1364-1369.
17. Lopez OL, Swihart AA, Becker JT, et al. Reliability of NINCDS-ADRDA clinical criteria for the diagnosis of Alzheimer's disease. Neurology 1990; 40: 1517-1522.
18. Baldereschi M, Amato MP, Nencini P, et al. Cross-national interrater agreement on the clinical diagnostic criteria for dementia. WHO-PRA Age-Associated Dementia Working Group, WHO-Program for Research on Aging, Health of Elderly Program. Neurology 1994; 44: 239-242.
19. Derouesne C. Dementia. Diagnostic problems. Therapie 1993; 48: 189-193.
20. Fischer P, Danielczyk W, Jellinger K, et al. Differential diagnosis of dementia diseases. A prospective clinical study with neuropathologic diagnostic verification. Nervenarzt 1991; 62: 408-414.
21. Galasko D, Hansen LA, Katzman R, et al. Clinical-neuropathological correlations in Alzheimer's disease and related dementias. Arch Neurol 1994; 51: 888-895.
22. Joachim CL, Morris JH, Selkoe DJ. Clinically diagnosed Alzheimer's disease: autopsy results in 150 cases. Ann Neurol 1988; 24: 50-56.
23. Kazee AM, Eskin TA, Lapham LW, et al. Clinicopathologic correlates in Alzheimer's disease: assessment of clinical and pathologic diagnostic criteria. Alzheimer Disease & Associated Disorders 1993; 7: 152-164.
24. Tierney MC, Fisher RH, Lewis AJ, et al. The NINCDS-ADRDA Work Group criteria for the clinical diagnosis of probable Alzheimer's disease: a clinicopathologic study of 57 cases. Neurology 1988; 38: 359-364.
25. Blacker D, Albert MS, Bassett SS, et al. Reliability and validity of NINCDS-ADRDA criteria for Alzheimer's disease. The National Institute of Mental Health Genetics Initiative. Arch Neurol 1994; 51: 1198-1204.

26. Gearing M, Mirra SS, Hedreen JC, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part X. Neuropathology confirmation of the clinical diagnosis of Alzheimer's disease . Neurology 1995; 45: 461-466.
27. Davis PC, Gearing M, Gray L, et al. The CERAD experience, Part VIII: Neuroimaging-neuropathology correlates of temporal lobe changes in Alzheimer's disease. Neurology 1995; 45: 178-179.
28. Jobst KA, Hindley NJ, King E, Smith AD. The diagnosis of Alzheimer's disease: a question of image? J Clin Psychiatry 1994; 55 (Suppl): 22-31.
29. Jobst KA, Smith AD, Szatmari M, et al. Rapidly progressing atrophy of medial temporal lobe in Alzheimer's disease. Lancet 1994; 343: 829-830.
30. Nagy Z, Jobst KA, Esiri MM, et al. Hippocampal pathology reflects memory deficit and brain imaging measurements in Alzheimer's disease: clinicopathologic correlations using three sets of pathologic diagnostic criteria. Dementia 1996; 7: 76-81.
31. Willmer J, Carruthers A, Guzman DA, et al. The usefulness of CT scanning in diagnosing dementia of the Alzheimer type. Can J Neurol Sci 1993; 20: 210-216.
32. Albert M, Naeser MA, Levine HL, Garvey AJ. Ventricular size in patients with presenile dementia of the Alzheimer's type. Arch Neurol 1984; 41: 1258-1263.
33. Kennard ML, Feldman H, Yamada T, Jefferies WA. Serum levels of the iron binding protein p97 are elevated in Alzheimer's disease. Nature Medicine 1996; 2: 1230-1235.
34. Albert MS. Cognitive and neurobiologic markers of early Alzheimer disease. Proc Natl Acad Sci U S A 1996; 93: 13547-13551.
35. Poirier J, Delisle MC, Quirion R, et al. Apolipoprotein E4 allele as a predictor of cholinergic deficits and treatment outcome in Alzheimer disease. Proc Natl Acad Sci U S A 1995; 92: 12260-12264.