Clinical research

Sensitivity and specificity of neuroimaging for the diagnosis of Alzheimer's disease

Daniel E. Wollman, MD, PhD; Isak Prohovnik, PhD

Alzheimer's disease (AD) is one of the most devastating and costly disorders affecting the aging population. Structural imaging (computed tomography [CT] and magnetic resonance imaging [MRI]) and functional imaging (single photon emission computed tomography [SPECT] and positron emission tomography [PET]) have been evaluated for their roles in the imaging diagnosis of AD. We have reviewed the recent literature to determine the capabilities of these neuroimaging techniques in comparison to current standards of clinical diagnosis. Our results indicate that there is wide variability in the accuracy of clinical assessments, in contrast to a more limited range of variability of the accuracy of neuroimaging measurements. These results suggest that neuroimaging may serve an adjunctive role in raising this lower bound of diagnostic accuracy. Furthermore, we suggest that neuroimaging should be considered: (i) when clinical expertise is insufficient; (ii) as a complement to specific likelihood ratios; and (iii) in specific types of patients, for whom clinical evaluation is inappropriate or inadequate.

Alzheimer's disease (AD) is one of the most devastating and costly disorders affecting the aging population. This disease has an estimated prevalence of up to 40% in those over age 80.[1] Its financial cost to society has been estimated at between $70 and $100 billion annually.[2] Currently approved therapies, arguably modest in effect, focus on symptomatic treatment.[3-5] Preventive strategies, on the other hand, remain elusive. Better understanding of this disorder, as well as the development of both preventive and improved symptomatic treatments, has been limited by difficulties encountered in clinical diagnosis and the lack of adequate quantitative biomarkers for the disease.

Clinical diagnosis depends on the definition of cognitive deficits and the separation of normal age-related decline from pathological deterioration. Because the normal range of variability of cognitive abilities among the aged is extremely large, it is difficult to quantify precise normative limits of the normal range. It is commonly understood that different levels of cognitive functioning are expected from a 90 year old than a 60 year old, or a university graduate versus an illiterate person. Instead, the clinical diagnosis of dementia usually relies on the characterization of intra-individual decline from premorbid level of functioning. Typically, however, firm quantitative data about premorbid status are lacking, and the diagnostic process relies instead on interviews at the time of symptomatic onset that attempt to characterize premorbid performance levels. This approach is limited in its accuracy, and suffers from possi-

Author affiliations: Department of Geriatrics, Mount Sinai School of Medicine, New York, NY, USA (Daniel E. Wollman); Departments of Psychiatry and Radiology, Mount Sinai School of Medicine, New York, NY, USA (Isak Prohovnik); Department of Diagnostic Radiology, Yale University, New Haven, CT, USA (Isak Prohovnik)

Address for correspondence: Isak Prohovnik, PhD, Bronx VAMC, 130 West Kingsbridge Road, Bronx, NY 10468, USA
(e-mail: isak.prohovnik@mssm.edu)

Clinical research

ble sources of diagnostic bias. In addition to these problems of isolating mild, initial AD from normal aging, the clinical diagnosis is sometimes ambiguous due to overlap of symptoms between AD and other dementing illnesses. To address these issues and improve diagnostic accuracy, we need to support the clinical diagnosis by laboratory markers. Many have been sought,[6-8] this article addresses one of the most promising and best documented, based on imaging of cerebral structure and function.

Several modalities as well as strategies (eg, quantitative versus qualitative) have been evaluated for their role in the imaging diagnosis of AD. Computed tomography (CT) and magnetic resonance imaging (MRI) have focused primarily on the structural changes observed in specific brain areas during the course of the disease. Studies evaluating the diagnosis of AD using these techniques are based on impressionistic (or interpretive) measures (eg, qualitative determination of atrophy) or more rigorous quantitative measures where linear or volumetric parameters are obtained from the imaging data. The mesial temporal lobe (MTL), especially the hippocampus, has emerged as the most sensitive area to examine for AD-related atrophy. Functional neuroimaging, such as single photon emission computed tomography (SPECT) or positron emission tomography (PET), typically measures cerebral perfusion or metabolism, reflecting alteration in cerebral function. These studies are also based on either qualitative impression or objective measured parameters. The area most sensitive to such functional deficits in AD is the inferior parietal cortex. There is a large body of evidence regarding the validity of both measures (hippocampal atrophy and parietal metabolic deficit) as markers of AD. The relationship between the two is obscure, and despite their promise, imaging findings lack compelling evidence for their diagnostic value. Recent diagnostic guidelines by the American Academy of Neurology[9] recommend:

- The National Institutes of Neurological, Communicative Disorders and Stroke (NINCDS)–Alzheimer's Disease and Related Disorders Association (ADRDA) for the diagnosis of probable AD or *Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition (DSM-III-R)*[10] criteria for dementia of Alzheimer's type (DAT) should be routinely used.
- Structural neuroimaging with either a noncontrast CT or MRI scan in the routine initial evaluation of patients with dementia is appropriate.
- Linear or volumetric MRI or CT measurement strategies for the diagnosis of AD are not recommended for routine use at this time.
- For patients with suspected dementia, SPECT cannot be recommended for routine use in either initial or differential diagnosis, as it has not been demonstrated to be superior to clinical criteria.
- PET imaging is not recommended for routine use in the diagnostic evaluation of dementia at this time.

The purpose of this article is to review the neuroimaging literature and suggest avenues of promising research for AD diagnostics. While we agree with the Academy's recommendation against routine neuroimaging in all cases, we believe that neuroimaging offers unique capabilities for this purpose, which may be extremely useful in some contexts. As mentioned recently by Hogan and McKeith,[11] the routine use of structural neuroimaging may be justifiable merely to detect the 5% of patients with clinically unsuspected structural lesions. In addition, we point out here a similarly infrequent, but important, need for functional imaging. Below we will analyze the literature with the aim of detecting these specific applications.

Methods

We performed a computerized search of the indexed medical literature (August 1998–August 2001) through Medline® using the following medical subheading (MeSH)

terms: *Alzheimer Disease/ AND Diagnostic Imaging/ AND Sensitivity/ AND Specificity/.* This search produced 13 citations that directly reported sensitivity and specificity in diagnosing or distinguishing AD from either normal or other diseased states (including non-AD dementia or other mental illness). We additionally searched the literature for data on the sensitivity and specificity of clinically based assessments, obtaining 9 studies for comparison. We categorized the results of each report according to the modality (eg, clinical, CT, MRI, SPECT, or PET), the strategy (measured or interpreted), and comparison group (normal controls or patients with other dementia types). Studies reporting sensitivity and specificity data for individual measures (eg, entorhinal cortex blood flow or sensorimotor cortex blood flow) were listed as separate entries. We constructed a database of these multiple criteria.

Early in the analysis, we encountered a complication in comparing clinical evaluation against neuroimaging. The ultimate diagnosis of AD is a neuropathological one. Clinical diagnosis is usually validated against clinical follow-up, or against postmortem neuropathological diagnosis. Neuroimaging studies have usually been validated against clinical diagnosis. This introduces difficulty into interpretation of the comparison, since there is a variable error associated with the clinical diagnosis. While it is not strictly proper to compare the accuracy of imaging diagnosis (against clinical diagnosis) with the accuracy of clinical diagnosis (against neuropathological findings), the comparison is heuristically useful.

Results

Clinical diagnosis

To provide a comparison for the accuracy of imaging data, and evaluate its cost–benefit characteristics, we first provide information on the accuracy of clinical diagnosis against postmortem neuropathology.

Simple screening measures, such as the Mini-Mental State Examination (MMSE), often provide good diagnostic accuracy. For example, Muller et al and Wahlund et al reported reasonable sensitivity and specificity values for the MMSE alone and in combination with a verbal recall test.[12,13] More informative results were obtained with standardized clinical measures when validated against neuropathological diagnosis. In Jobst et al, 200 affected cases were compared with normal controls by standardized clinical measures, and then validated with histopathologic diagnosis.[14] Using NINCDS possible or probable AD criteria, Jobst et al reported a maximum sensitivity of 96%, with associated specificity of 61%. In the same study, the use of *DSM-III-R* criteria applied to the same study groups resulted in a sensitivity of 51%, and specificity of 97%. Other authors, noted in *Table I*, obtained similar results.[15-21] Overall, the range of sensitivity of clinical diagnosis was 39% to 98%, and the range of specificity was 33% to 100%. There was a significant negative correlation (*r*=-0.79, *P*=0.01) between sensitivity and specificity, as expected, reflecting the necessary tradeoff. Thus, for instance, to achieve a specificity

	No. AD subjects	No. controls	Clinical criteria	Sensitivity	Specificity	Neuropathological criteria
Boller et al,[20] 1989	39	15 other	NINCDS probable AD	0.95	0.33	Other
Hoffman et al,[15] 2000	9	13 other	NINCDS probable AD	0.64	0.88	CERAD
Jobst et al,[14] 1998	80	38 other	NINCDS probable AD	0.49	1.00	CERAD
Jobst et al,[14] 1998	80	38 other	*DSM-III-R*	0.51	0.97	CERAD
Kazee et al,[19] 1993	94	29 normal	NINCDS probable AD	0.98	0.69	Other
Lim et al,[16] 1999	84	36 other	NINCDS probable AD	0.83	0.55	CERAD
Massoud et al,[17] 2000	25	36 other	NINCDS probable AD	0.86	0.50	CERAD
Nagy et al,[18] 1998	46	27 other	NINCDS probable AD	0.41	1.00	CERAD
Nagy et al,[18] 1998	46	27 other	*DSM-III-R*	0.39	0.96	CERAD
Tierney et al,[21] 1988	22	35 other	NINCDS probable AD	0.86	0.89	Other

Table I. Sensitivity and specificity of clinical measurements. AD, Alzheimer's disease; CERAD, CERAD (Consortium to Establish a Registry for Alzheimer Disease) probable or definite AD (neuropathology); Other, other neuropathological review; *DSM-III-R, Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition*; NINCDS, National Institutes of Neurological, Communicative Disorders and Stroke.

greater than 80%, four out of five studies had to settle for sensitivity lower than 70%. This correlation is depicted in *Figure 1*.

A number of studies used the criteria "NINCDS possible or probable AD" or other nonstandard clinical measures (data not shown).[14-18,20,22-27] While clinical diagnosis is often used to validate imaging findings, and neuro-pathological diagnosis is the overall "gold standard," and despite the existence of modern standardized criteria, the application of these standards should not be considered free of ambiguity. A good example is provided by Hoffman et al.[15] The clinical NINCDS criteria allow the definition of probable or possible AD, reflecting different degrees of confidence. The Consortium to Establish a Registry for Alzheimer Disease (CERAD) pathological criteria allow the finding of "pure" AD or AD in addition to other pathology. These authors reported the sensitivity and specificity values for these four groups of patients (probable/possible clinical AD validated by pure or nonpure pathological AD diagnosis). For these four groups, sensitivity varied from 63% to 79%, and specificity from 88% to 100%. Discussion of these issues is outside the scope of this article, but such distinctions add noise to the data reviewed here and should be kept in mind when evaluating the material. When finding such multiple data, we made the choice of reporting the values for probable AD only, but did not require "pure" pathological AD.
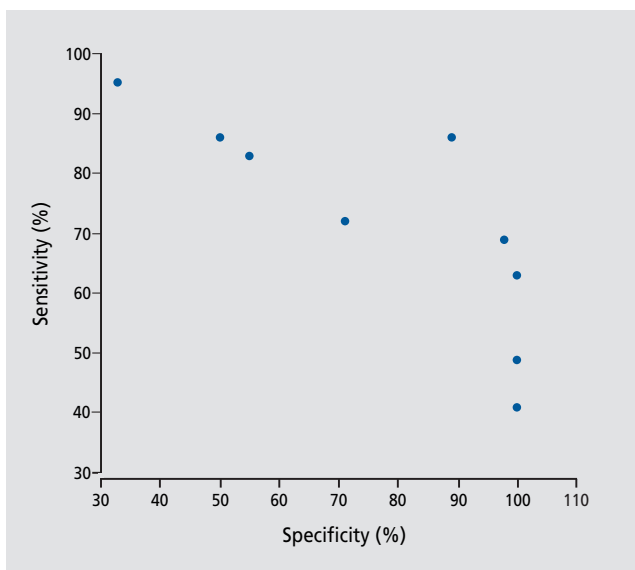

Figure 1. Sensitivity and specificity of clinical diagnosis against neuro-pathological diagnosis.

Computed axial tomography

Two fairly large studies using CT scanning techniques were available for review. Jobst et al[14] reported diagnostic accuracy based on the measured width of the MTL, ultimately compared with neuropathology. This technique resulted in sensitivity of 85% and specificity of 78%. In the same study, they combined this measurement with an impressionistic measure using SPECT to visualize decreased parietotemporal perfusion. The addition of the impressionistic measure worsened sensitivity to 80%, but resulted in improved specificity of 93%. A second study by Denihan et al[28] compared a series of patients with AD versus non-AD controls with vascular dementia, depression, or paraphrenia. This study, also based on the measured width of the MTL, resulted in sensitivity of 75% and specificity of 90%. These results are summarized in *Table II*.

Magnetic resonance imaging

The application of MRI to the diagnosis of AD is currently in active research and development, and shows significant promise. Modern MRI can also provide functional measures, such as cerebral blood volume, blood flow, and velocity. Some of these measures show great potential, similar to PET and SPECT (eg, sensitivity of 93% and specificity of 94% for parietotemporal blood volume relative to cerebellum[29]), but this review is only concerned with structural MRI, as this represents the most well-studied technique of this imaging modality. Volumetric analyses of MTL structures show good discrimination from normal aging. However, it is not yet established which single measure (if any) is best, and the literature contains references to several measures. In our search, four studies using MRI reported a total of 22 measurements; the most relevant to the diagnosis of AD are included in *Table III*. In comparing subjects with AD versus normal controls, the best sensitivity/specificity measurements (greater than 80% each) were achieved utilizing quantitative measurement of hippocampal volume (95%/92%),[30] entorhinal cortex volume (90%/94%),[31] and MTL volume (88%/96%).[13] One study[13] reported qualitative impression of MTL volume in combination with the MMSE, resulting in sensitivity and specificity of 93% and 98%, respectively. This same study also examined the addition of MRI volumetry to MMSE in distinguishing AD from other dementia. Sensitivity and specificity in this case

was 68% and 53%, respectively. Impression of atrophy added little to sensitivity and specificity (78% and 64%) over objective measurement. These data are not included in the numerical analysis or *Figure 2*. On observation, however, these data indicate that the addition of an imaging measurement adds little to an already relatively high sensitivity for clinical assessment in the case of AD versus normal controls.

Positron emission tomography

Table IV illustrates the results of PET studies. The most notable is the report by Silverman et al,[2] which combined results of 284 PET studies, including 138 with histopathologic diagnoses and the others with 2 years' clinical follow-up. The scans were interpreted by nuclear medicine physicians and classified into profiles. AD was identified (blind to clinical information) with a sensitivity of 94% and specificity of 73%. Similarly, Hoffman[15] qualitatively examined parietotemporal (PTC) hypometabolism achieving sensitivity and specificity of 93% and 63%, respectively. There were two studies that examined the distinction of AD from dementia with Lewy bodies (DLB).[32,33] These studies achieved diagnostic sensitivity of 86% and 92%. The data from these two studies are not included in the numerical analysis as they represent a fundamentally different measurement than that used in the diagnosis of AD and more appropriately represent a measurement that distinguishes subjects with DLB.
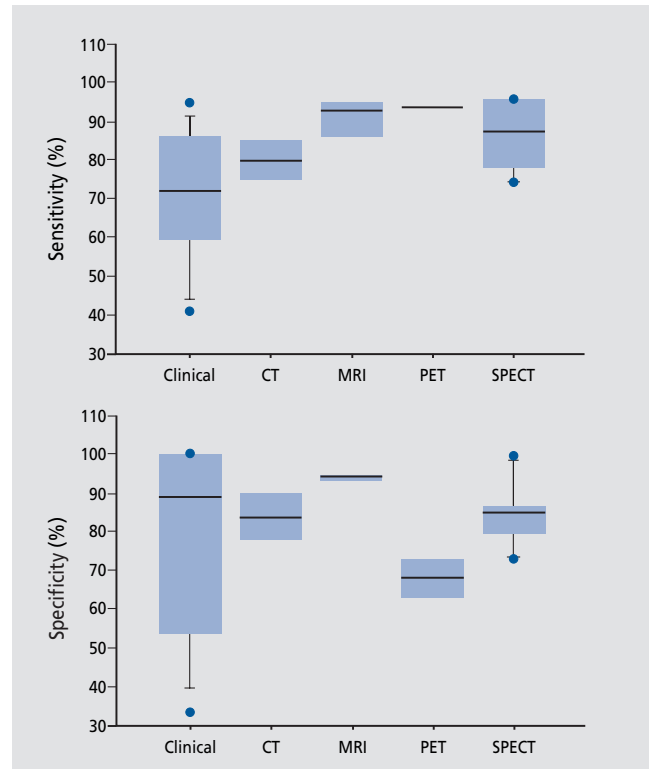

Figure 2. Summary of sensitivity and specificity of clinical and imaging modalities. The box plots show the distribution of values in each category by indicating the 10th, 25th, 50th, 75th, and 90th percentiles. Values below 10% or above 90% are depicted as individual points. CT, computed tomography; MRI, magnetic resonance imaging; PET, positron emission tomography; SPECT, single photon emission computed tomography.

	No. AD subjects	No. controls	Measurement/impression	Sensitivity	Specificity	Diagnostic criteria
Jobst et al,[14] 1998	200	119 normal	(M) MTL width	0.85	0.78	CERAD
Jobst et al,[14] 1998	200	119 normal	(M) MTL width (I) PTC rCBF	0.80	0.93	CERAD
Denihan et al,[28] 2000	60	40 other	(M) MTL width	0.75	0.90	NINCDS

Table II. Sensitivity and specificity of computed tomography measures. AD, Alzheimer's disease; CERAD, CERAD (Consortium to Establish a Registry for Alzheimer Disease) probable or definite AD (neuropathology); NINCDS, NINCDS (National Institutes of Neurological, Communicative Disorders and Stroke) probable AD (clinical); MTL, mesial temporal lobe; PTC, parietotemporal cortex; rCBF, regional cerebral blood flow; M, measurement; I, impression.

	No. AD subjects	No. controls	Measurement	Sensitivity	Specificity	Diagnostic criteria
Harris et al,[29] 1998	27	18 normal	PTC rCBV	0.93	0.94	NINCDS
Juottonen et al,[31] 1999	55	83 other	Hippocampal volume	0.90	0.94	NINCDS
Wahlund et al,[13] 2000	41	67 normal	MTL volume + MMSE	0.88	0.96	NINCDS possible
Golebiowski et al,[30] 1999	50	25 normal	Hippocampal volume	0.95	0.92	NINCDS other

Table III. Sensitivity and specificity of magnetic resonance imaging measures. AD, Alzheimer's disease; NINCDS, NINCDS (National Institutes of Neurological, Communicative Disorders and Stroke) probable AD (clinical); NINCDS possible, NINCDS possible or probable AD (clinical); NINCDS other, unspecified NINCDS criteria (clinical); rCBV, regional cerebral blood volume; PTC, parietotemporal cortex; MTL, mesial temporal lobe.

Single photon emission computed tomography

The widest variation in diagnostic accuracy overall was apparent in the studies using SPECT. Seven studies reported a total of 35 measurements; the most relevant to the diagnosis of AD are included in *Table V*.[14,29,34-37] The best sensitivity/specificity in distinguishing subjects with AD versus normal controls reached 96%/87%, by calculating a discriminant function based on regional cerebral blood flow (rCBF) of multiple brain regions.[34] Impressionistic studies of decreased parietotemporal blood flow achieved a maximal sensitivity/specificity of 89%/80%.[14] One impressionistic study compared subjects with AD with DLB, resulting in sensitivity/specificity as low as 65%/87%.[35] Sjogren et al[36] examined the utility of quantitative SPECT in several dementia subtypes. In each of the reported measurements, specificity was arbitrarily set at 85%. In subjects with frontotemporal dementia, maximal sensitivity/specificity achieved was 81%/85%, examining the rCBF of the superior frontal gyrus. In early-stage AD, measurement of rCBF of the MTL results in sensitivity of 85%. This measurement improved to 96% for subjects with late-stage AD. Interestingly, measurement of rCBF of the PTC results in sensitivity of 90% for dementia associated with subcortical white matter disease. Measurements of blood flow in other brain structures such as white matter, hippocampus, or structures not affected in the particular dementia under study, resulted in diagnostic sensitivity often far below 80%, and are not included in this review.

Discussion

Neuroimaging is fairly expensive, complex, and requires specialized facilities and expertise that may not always be easily available. Its routine use thus requires rational examination of cost–benefit considerations. For the purpose of AD diagnosis, the recent Academy of Neurology report[9] concludes—and this review supports—that clinical diagnosis can be quite effective. In the most skilled hands and under favorable conditions, the accuracy of clinical diagnosis can be very high, as confirmed by histopathologic diagnosis. Sensitivity and specificity data of 85% or better are commonly reported. Therefore, the routine use of neuroimaging was not recommended by the recent Academy report, nor does it appear justified by our data. While it may be premature to recommend neuroimaging in all evaluations of dementia, there is a clear role for neuroimaging in certain circumstances and, as such, neuroimaging may play a role in offering true, objective determinations of the disease state. We agree with the conclusion that neuroimaging offers, at best, the same level of diagnostic accuracy as *expert* clinical assessment. Thus, from a cost-effectiveness viewpoint, neuroimaging currently offers no additional benefit over intensive, clinically based assessments. One must consider, however, that clinical assessment requires a level of expertise, as well as optimal circumstances for test administration that may not always be possible. Additionally, there are confounding circumstances compromising the validity and accuracy of clinical assessment. Three sets of observations suggest that neuroimaging should be considered, and offers favorable cost–benefit ratio, in some circumstances. These are: (i) when clinical expertise is insufficient; (ii) as a complement to specific likelihood ratios; and (iii) in specific types of patients. These circumstances are discussed below.

Clinical expertise

An overall summary of this review is provided in *Figure 2*. Note that all clinical diagnoses were evaluated against histopathology, whereas some imaging findings were validated by clinical diagnosis. Nevertheless, the data suggest the following observations. First, the specificity of clinical diagnosis may be better than its sensitivity (77±26% vs 72±18%, NS in this sample). The mean specificity of clinical diagnosis compares favorably with the values offered by neuroimaging, but mean sensitivity of clinical diagnosis is lower. More striking, however, are the differences in variance. By any measure of dis-

	No. AD subjects	No. controls	Impression	Sensitivity	Specificity	Diagnostic criteria
Hoffman et al,[15] 2000	14	8 other	PTC hypometabolism	0.93	0.63	CERAD
Silverman et al,[2] 2001	97	41 other	Cerebral hypometabolism	0.94	0.73	Other

Table IV. Sensitivity and specificity of positron emission tomography measures. AD, Alzheimer's disease; NINCDS, NINCDS (National Institutes of Neurological, Communicative Disorders and Stroke) probable AD (clinical); CERAD, CERAD (Consortium to Establish a Registry for Alzheimer Disease) probable or definite AD (neuropathology); Other, other neuropathological criteria; DLB, dementia with Lewy bodies; CMRglu, cerebral metabolic rate of glucose; PTC, parietotemporal cortex.

person, clinical diagnosis accuracy is far more variable in this material than the accuracy of any imaging method. The range of sensitivity of clinical diagnosis is 34% to 95%, and the range of specificity 33% to 100%. Clearly, these values range from perfect to unacceptable. This variability of clinical diagnostic accuracy can probably be attributed to several factors. It includes the relatively large number of studies reviewed, characteristics of patient and control samples, limited reproducibility of clinical ratings, and perhaps even different neuropathological procedures. Another source of variance may be the result of imperfect clinical criteria. Both NINCDS and the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)*[38] criteria sets contain features dependent on the skill of the clinician, as well as features requiring qualitative determination, possibly rendering the criteria subject to variable interpretation. In the NINCDS criteria, a diagnosis of "probable AD" requires the establishment of dementia by (i) MMSE or Blessed Dementia Scale; and (ii) confirmatory neuropsychological testing. In addition, there must be a "progressive worsening of memory and other cognitive functions." While the former features are for the most part objective measures, the latter feature is not specified in detail and might be interpreted in a subjective manner. The alternative criteria delineated in *DSM-IV* do not require objective testing, thus permitting a diagnosis of AD solely on subjective grounds. Thus, a clinician employing *DSM-IV* criteria might diagnose AD solely from the patient's history without seeking confirmatory, objective testing. This approach limits the standardization of diagnosis and depends heavily on the diagnostician's skills.

Indeed, we believe that the main factor responsible for the variability in clinical diagnosis is the individual skill, experience, and expertise of the diagnostician. Training, experience, and insight vary substantially, and probably affect accuracy. Further, the clinical assessment of AD occurs primarily in two settings: (i) primary care screening; and (ii) consultative evaluation of memory or cognitive complaints. Consultative evaluations by memory specialists are more likely to employ formal testing. In either setting, the intensity of the evaluation is often physician-dependent, though clear guidelines exist suggesting appropriate testing and criteria to be used in the diagnosis.

In contrast, the range of accuracies for imaging findings is much more limited, typically of the order of 10% to 15%. Imaging procedures are often well standardized, and commonly performed by technicians as a matter of fixed routine. While the interpretation of imaging results is often a matter of skill and expertise, much like clinical diagnosis,[14] AD diagnosis has matured to the extent that many papers report quantitative, measured results, rather than an interpretation of patterns. Thus, much of the variance is removed.

Thus, while the best clinicians under favorable circumstances achieve near-perfect diagnostic accuracy (at least with respect to sensitivity), some clinical evaluations suffer much lower accuracy. Neuroimaging procedures, especially with measured (rather than interpreted) outcomes, are much more consistent and much less dependent on individual skills. It appears, thus, that neuroimaging procedures can be of significant value in circumstances where an expert clinician is not readily available.

	No. AD subjects	No. controls	Measurement/impression	Sensitivity	Specificity	Diagnostic criteria
Bonte et al,[37] 1997	37	16 other	(I) rCBF	0.86	0.73	Other
Harris et al,[29] 1998	19	18 normal	(I) rCBF PTC	0.74	1.00	NINCDS
Jobst et al,[14] 1998	80	38 other	(I) rCBF PTC	0.89	0.80	NINCDS
Lobotesis et al,[35] 2001	50	20 normal	(M) rCBF PTC	0.78	0.85	NINCDS other
Sjogren et al,[36] 2000	25 (late)	28 normal	(M) rCBF MTL	0.96	0.85	NINCDS
Sjogren et al,[36] 2000	25 (late)	28 normal	(M) rCBF PTC	0.72	0.85	NINCDS
Tsolaki et al,[34] 2001	117	41 other	(M) rCBF MTL	0.96	0.87	NINCDS possible
Sjogren et al,[36] 2000	27 (early)	28 normal	(M) rCBF PTC	0.82	0.85	NINCDS
Sjogren et al,[36] 2000	27 (early)	28 normal	(M) rCBF MTL	0.85	0.85	NINCDS

Table V. Sensitivity and specificity of single photon emission computed tomography measures. AD, Alzheimer's disease; NINCDS, NINCDS (National Institutes of Neurological, Communicative Disorders and Stroke) probable AD (clinical); Other, other neuropathological analysis; NINCDS other, NINCDS unspecified (clinical); NINCDS possible, NINCDS possible AD (clinical); rCBF, regional cerebral blood flow; PTC, parietotemporal cortex; MTL, mesial temporal lobe; M, measurement; I, impression.

Clinical research

Complementing likelihood ratios

As demonstrated earlier in *Figure 1*, clinical diagnosis usually involves a tradeoff between sensitivity and specificity, even when using standardized clinical scales. Partly as a function of the scales used, partly depending on explicit or implicit cutoff selection, and partly due to imperfect reliability, clinical diagnosis commonly offers either good sensitivity or good specificity, but not both. On average, specificity is better than sensitivity *(Figure 2)*. Further, circumstances tend to emphasize one or the other. For example, if treatment is toxic or difficult to institute, specificity should probably be maximized. On the other hand, if treatment is benign, but needs to be initiated in the early stages of the disease, sensitivity is more important. This is exemplified most clearly by recent suggestions of the relationship between dementia and statin use[39] or suggestion of early cholinesterase use in mild cognitive impairment (MCI).[40] Neuroimaging may help distinguish those individuals with MCI likely to develop AD.[41]

Studies that compared both clinical diagnosis and imaging findings to eventual neuropathological diagnosis are especially noteworthy. Hoffman et al,[15] for example, achieved sensitivity/specificity values of 63%/100% for the clinical diagnosis of probable AD in a small sample; the corresponding values for the parietotemporal metabolic deficit were 93%/63%. In this case, therefore, imaging was not superior overall to clinical examination. However, because imaging appeared more sensitive and clinical diagnosis, more specific, overall accuracy could be substantially improved if the two were combined. Unfortunately, the sensitivity advantage of imaging is not always reproduced. Furthermore, while current state of knowledge is not definitive, there are indications that the various imaging modalities are not identical with respect to predictive properties. For example, the current data suggest that PET offers high sensitivity but lower specificity. It would therefore be more appropriate in circumstances where maximal sensitivity is sought. Thus, given more precise knowledge about the predictive properties of various clinical and imaging methods, one could complement a sensitive clinical assessment with a specific imaging procedure, and vice versa, thus maximizing diagnostic yield.

Patient characteristics

Clinical diagnosis of AD is easier at advanced stages of the disease; it can be very difficult during the insidious onset. It is likely that neuroimaging suffers from the same limitation, although possibly not to the same extent.[36] Thus, neuroimaging may be especially beneficial in the very early stages. Moreover, once presymptomatic treatment trials begin, it is likely that neuroimaging may be of unique value in identifying patients likely to convert to symptomatic status in the future.

In addition to the severity and duration of the disease, other confounding factors in the clinical diagnosis of AD include variables such as the patient's age, level of education, and native language. Most patients included in research protocols are relatively young, whereas most patients in the population are older. It is not yet known how clinical diagnostic accuracy varies across the age span, and in the presence of comorbidities more prevalent in the older age range. Education has been shown to affect the incidence of the disease and/or the likelihood of being diagnosed.[42] Certainly, cognitive performance, as measured by screening instruments like the MMSE, is affected by age as well as education. Those individuals with advanced education may be characterized as normal on initial evaluation, only to be seen in later courses of the disease when symptoms are more apparent and the dementia more severe.[43,44] We have previously documented that patients matched for current clinical dementia severity demonstrate different degrees of brain damage as measured by imaging procedures.[45] Finally, existing neuropsychological testing in other languages may not be available (or validated) for non–native English-speaking subjects. The use of existing English-based tests in non–native English-speaking subjects may be inaccurate or insensitive in these circumstances.[46]

Fundamentally, the onset of AD consists of a decline from premorbid level of functioning. This premorbid level, the "normal" level, is extremely variable in the normal population across age, language skills, educational and occupational background, etc. For this reason, it is difficult to clinically assess decline in the absence of strong documentation of premorbid functioning. Neuroimaging may offer this capability. We have previously shown that the parietotemporal perfusion deficit is strongly related to this decline, even more strongly than to current dementia severity.[45] Thus, the fundamental measurement of decline from premorbid levels may be possible with functional neuroimaging. If confirmed in future studies, this capability may overcome all factors currently confounding clinical diagnosis: regard-

less of the patient's language skills, educational background, or age, we may be able to define how much their brain function has declined from what was, for each individual patient, normal levels. This decline may well be a better predictor of progression or medication response than current clinical symptomatology.

Conclusion

We have reviewed the recent literature on neuroimaging diagnosis of AD. As in any conclusions based on a literature review or meta-analysis, the possibility of a publication bias must be considered. It is possible that unsuccessful imaging studies (ie, those reporting low diagnostic accuracy) are not published, due to reservations by authors or editors. It is also possible that imaging papers tend to be submitted to specific journals, with publication policies different from those of other, more purely clinical, journals. Finally, some papers may have been published in journals not indexed by Medline. Thus, further consideration of our conclusions must be bound by the nature of the material and its limitations. Our interpretation of this literature offers two main conclusions. First, that the variability of diagnostic accuracy is considerably lower than that of clinical diagnosis. In particular, while neuroimaging cannot improve the best clinical diagnosis findings (which are close to 100%), the lowest accuracies reported for imaging are considerably higher than the lowest accuracies reported for clinical diagnosis *(Figure 2)*. Thus, imaging can serve to significantly improve the lower bounds of diagnostic accuracy.

Second, we propose that imaging adds unique information to the diagnostic process that may not be available by any other methods. This information may be especially pertinent in certain clinical situations, discussed above. Both clinical criteria and imaging procedures are continuously evolving, and they need to continue to be used together for further evaluation. While MRI appears to be superior overall in this material *(Figure 2)*, the current work was not designed to compare the relative merits of various imaging modalities. Studies that employ more than one imaging modality are rare but useful, and more need to be conducted. For example, De Santi et al[41] compared PET-derived glucose metabolism and MRI-derived volumetric measures in temporal lobe structures. They concluded, overall, that neocortical (middle and superior temporal gyrus) measures were more accurate than hippocampal structures, and that functional PET measures were superior to MRI findings in discriminating AD from normal controls. Like others, these authors showed that the relative classification merit of various structures in and around the hippocampal formation is variable, and there is still no agreement on the most informative structures. Because sensitivity/specificity values were not rigorously reported, we did not include this study in our tables and figures, but clearly more studies of this type are necessary. ❏

REFERENCES

1. Evans DA, Funkenstein HH, Albert MS, et al. Prevalence of Alzheimer's disease in a community population of older persons: higher than previously reported. *JAMA*. 1989;262:2551-2556.
2. Silverman DHS, Small GW, Chang CY, et al. Positron emission tomography in evaluation of dementia: regional brain metabolism and long-term outcome. *JAMA*. 2001;286:2120-2127.
3. Davis KL, Thal LJ, Gamzu ER, et al. A double-blind, placebo-controlled multicenter study of tacrine for Alzheimer's disease. *N Engl J Med*. 1992;327:1253-1259.
4. Rogers SL, Friedhoff LT. The efficacy and safety of donepezil in patients with Alzheimer's disease: results of a US multicentre, randomized, double-blind, placebo-controlled trial. *Dementia*. 1996;7:293-303.
5. Raskind MA, Peskind ER, Wessel T, et al. Galantamine in AD—a 6-month randomized, placebo-controlled trial with a 6-month extension. *Neurology*. 2000;54:2261-2268.
6. Mayeux R, Saunders AM, Shea S, et al. Utility of the apolipoprotein E genotype in the diagnosis of Alzheimer's disease. *N Engl J Med*. 1998;338:506-511.

7. Adalsteinsson E, Sullivan EV, Kleinhans N, Spielman DM, Pfefferbaum A. Longitudinal decline of the neuronal marker *N*-acetyl aspartate in Alzheimer's disease. *Lancet*. 2000;355:1696-1697.
8. Andreasen N, Minthon L, Davidsson P, et al. Evaluation of CSF-tau and CSF-AB42 as diagnostic markers for Alzheimer disease in clinical practice. *Arch Neurol*. 2001;58:373-379.
9. Knopman DS, DeKosky ST, Cummings JL, et al. Practice parameter: diagnosis of dementia (an evidence-based review). Report of the quality standards subcommittee of the American Academy of Neurology. *Neurology*. 2001;56:1143-1153.
10. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 3rd ed, revised. Washington, DC: American Psychiatric Association; 1987.
11. Hogan DB, McKeith IG. Of MCI and dementia: improving diagnosis and treatment. *Neurology*. 2001;56:1131-1132.
12. Muller H, Moller HJ, Stippel A, et al. SPECT patterns in probable Alzheimer's disease. *Eur Arch Psychiatry Clin Neurosci*. 1999;249:190-196.
13. Wahlund LO, Julin P, Johansson SE, Scheltens P. Visual rating and volumetry of the medial temporal lobe on magnetic resonance imaging in dementia: a comparative study. *J Neurol Neurosurg Psychiatry*. 2000;69:630-635.

Clinical research

Sensibilidad y especificidad de las neuroimágenes en el diagnóstico de la Enfermedad de Alzheimer

La Enfermedad de Alzheimer (EA) es uno de los trastornos más devastadores y de alto costo que afecta a la población senescente. Por su papel en el diagnóstico imagenológico de la EA se ha evaluado la imagenología estructural (tomografía computada [TC] y resonancia nuclear magnética [RNM]) y la imagenología funcional (tomografía computada por emisión de fotón único [SPECT] y tomografía por emisión de positrones [PET]). Se ha revisado la literatura reciente para determinar las capacidades de estas técnicas de neuroimágenes en comparación con los estándares actuales del diagnóstico clínico. Nuestros resultados señalan que existe una amplia variabilidad en la exactitud de las evaluaciones clínicas en contraste con un rango más limitado de variabilidad de la exactitud de las mediciones de las neuroimágenes. Estos resultados sugieren que la neuroimagenología puede servir como método complementario para elevar este menor límite de la exactitud diagnóstica. Por último, se sugiere que las neuroimágenes se deben tener en consideración (i) cuando la experiencia clínica es insuficiente, (ii) como complemento para frecuencias específicas de probabilidad y (iii) en algunos pacientes especiales, para quienes la evaluación clínica es inapropiada o no es suficiente.

Sensibilité et spécificité de la neuro-imagerie pour le diagnostic de la maladie d'Alzheimer

La maladie d'Alzheimer (MA) est l'un des troubles les plus dévastateurs et coûteux touchant la population âgée. L'imagerie structurale (tomodensitométrie et IRM) et l'imagerie fonctionnelle (SPECT et PET) ont été évaluées pour leur rôle dans le diagnostic de la MA. Nous avons passé en revue la littérature récente pour déterminer les capacités de ces techniques de neuro-imagerie comparées aux méthodes de référence actuelles de diagnostic clinique. Nos résultats montrent qu'il y a une grande variabilité dans la fidélité des évaluations cliniques par opposition à une moins grande variabilité de l'exactitude des mesures réalisées grâce à la neuro-imagerie. Ces résultats suggèrent que la neuro-imagerie peut jouer un rôle d'appoint pour améliorer la limite inférieure de l'exactitude du diagnostic. De plus, nous suggérons que la neuro-imagerie devrait être envisagée : (i) lorsque l'expertise clinique est insuffisante ; (ii) comme complément aux rapports de probabilité spécifiques ; et (iii) pour des types particuliers de patients pour lesquels une évaluation clinique est inappropriée ou insuffisante.

14. Jobst KA, Barnetson LP, Shepstone BJ, et al. Accurate prediction of histologically confirmed Alzheimer's disease and the differential diagnosis of dementia: the use of NINCDS-ADRDA and *DSM-III-R* criteria, SPECT, X-ray CT, and Apo E4 in medial temporal lobe dementias. *Int Psychogeriatr*. 1998;10:271-302.

15. Hoffman JM, Welsh-Bohmer KA, Hanson M, et al. FDG PET imaging in patients with pathologically verified dementia. *J Nucl Med*. 2000;41:1920-1928.

16. Lim A, Tsuang D, Kukull W, et al. Clinico-neuropathological correlation of Alzheimer's disease in a community-based case series. *J Am Geriatr Soc*. 1999;47:564-569.

17. Massoud F, Devi G, Moroney JT, et al. The role of laboratory studies and neuroimaging in the diagnosis of dementia: a clinicopathologic study. *J Am Geriatr Soc*. 2000;48:1204-1210.

18. Nagy Z, Esiri MM, Hindley NJ, et al. Accuracy of clinical operational diagnostic criteria for Alzheimer's disease in relation to different pathological diagnostic protocols. *Dement Geriatr Cogn Disord*. 1998;9:219-226.

19. Kazee AM, Eskin TA, Lapham LW, et al. Clinicopathologic correlates in Alzheimer disease: assessment of clinical and pathologic diagnostic criteria. *Alzheimer Dis Assoc Disord*. 1993;7:152-164.

20. Boller F, Lopez OL, Moosy J. Diagnosis of dementia: clinicopathologic correlations. *Neurology*. 1989;39:76-70.

21. 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.

22. Molsa PK, Paljarvi L, Rinne JO, Rinne UK, Sako E. Validity of clinical diagnosis in dementia: a prospective clinicopathological study. *J Neurol Neurosurg Psychiatry*. 1985;48:1085-1090.

23. Todorov AB, Go RCP, Constantinidis J, Elston RC. Specificity of the clinical diagnosis of dementia. *J Neurol Sci*. 1975;26:81-98.

24. Kukull WA, Larson EB, Reifler BV, Lampe TH, Yerby MS, Hughes JP. The validity of 3 clinical diagnostic criteria for Alzheimer's disease. *Neurology*. 1990;40:1364-1369.

25. Wade JPH, Mirsen TR, Hachinski VC, et al. The clinical diagnosis of Alzheimer's disease. *Arch Neurol*. 1987;44:24-29.

26. Huff FJ, Becker JT, Belle SH, et al. Cognitive deficits and clinical diagnosis of Alzheimer's disease. *Neurology*. 1987;37:1119-1124.

27. Morris JC, McKeel DW, Fulling K, Torack R, Berg L. Validation of clinical diagnostic criteria for Alzheimer's disease. *Ann Neurol*. 1988;24:17-22.

28. Denihan A, Wilson G, Cunningham C, Coakley D, Lawlor BA. CT measurement of medial temporal lobe atrophy in Alzheimer's disease, vascular dementia, depression and paraphrenia. *Int J Geriatr Psychiatry*. 2000;15:306-312.

29. Harris GJ, Lewis RF, Satlin A, et al. Dynamic susceptibility contrast MR imaging of regional cerebral blood volume in Alzheimer disease: a promising alternative to nuclear medicine. *AJNR Am J Neuroradiol*. 1998;19:1727-1732.

30. Golebiowski M, Barcikowska M, Pfeffer A. Magnetic resonance imaging–based hippocampal volumetry in patients with dementia of the Alzheimer type. *Dement Geriatr Cogn Disord*. 1999;10:284-288.

31. Juottonen K, Laakso MP, Partanen K, Soininen H. Comparative MR analysis of the entorhinal cortex and hippocampus in diagnosing Alzheimer disease. *AJNR Am J Neuroradiol*. 1999;20:139-144.

32. Higuchi M, Tashiro M, Arai H, et al. Glucose hypometabolism and neuropathological correlates in brains of dementia with Lewy bodies. *Exp Neurol*. 2000;162:247-256.

33. Ishii K, Sasaki M, Yamaji S, et al. Regional cerebral glucose metabolism in dementia with Lewy bodies and Alzheimer's disease. *Neurology*. 1998;51:125-130.

34. Tsolaki M, Sakka V, Gerasimou G, et al. Correlation of rCBF (SPECT), CSF tau, and cognitive function in patients with dementia of the Alzheimer's type, other types of dementia, and control subjects. *Am J Alzheimers Dis Other Dement*. 2001;16:21-31.

35. Lobotesis K, Fenwick JD, Phipps A, et al. Occipital hypoperfusion on SPECT in dementia with Lewy bodies but not AD. *Neurology*. 2001;56:643-649.

36. Sjogren M, Gustafson L, Wikkelso C, Wallin A. Frontotemporal dementia can be distinguished from Alzheimer's disease and subcortical white matter dementia by an anterior-to-posterior rCBF SPET ratio. *Dement Geriatr Cogn Disord*. 2000;11:275-285.

37. Bonte FJ, Weiner MF, Bigio EH, White CL. Brain blood flow in the dementias: SPECT with histopathologic correlation in 54 patients. *Radiology*. 1997;202:793-797.

38. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: American Psychiatric Association; 1994.

39. Jick H, Zornberg G.L, Jick SS, Seshadri S, Drachman DA. Statins and the risk of dementia. *Lancet*. 2000;356:1627¯1631.

40. Petersen RC, Doody R, Kurz A, et al. Current concepts in mild cognitive impairment. *Arch Neurol*. 2001;58:1985-1992.

41. De Santi S, de Leon MJ, Rusinek H, et al. Hippocampal formation glucose metabolism and volume losses in MCI and AD. *Neurobiol Aging*. 2001;22:529-539.

42. Friedland RP. Epidemiology, education, and the ecology of Alzheimer's disease. *Neurology*. 1993;43:246-249.

43. Crum RM, Anthony JC, Bassett SS, Folstein MF. Population-based norms for the Mini-Mental State Examination by age and education level. *JAMA*. 1993;269:2386-2391.

44. Ganguli M, Ratcliff G, Huff FJ, et al. Effects of age, gender, and education on cognitive tests in a rural elderly community sample: norms from the Monongahela Valley Independent Elders Survey. *Neuroepidemiology*. 1991;10:42-52.

45. Keilp JG, Prohovnik I. Intellectual decline predicts the parietal perfusion deficit in Alzheimer's disease. *J Nucl Med*. 1995;36:1347-1354.

46. Pandav R, Fillenbaum G, Ratcliff G, Dodge H, Ganguli M. Sensitivity and specificity of cognitive and functional screening instruments for dementia: the Indo-US Dementia Epidemiology Study. *J Am Geriatr Soc*. 2002;50:554-561.