Measurement of progression in Alzheimer’s disease: a clinician’s perspective

Douglas J. Gelb*†

Department of Neurology, University of Michigan Medical School, Ann Arbor, Michigan 48109-0316, U.S.A.

SUMMARY

Patients with Alzheimer’s disease (AD) and their families must confront two fundamental truths. First, AD is a uniformly progressive disease that ultimately results in debilitating cognitive impairment. Second, although there is now evidence that some medications may produce transient improvement or possibly even slowing of disease progression, there is currently no way to halt the progression of AD. Consequently, patients and their families consistently ask the following questions: 1. What new management issues can be anticipated, and when? 2. What clinical developments are atypical and merit evaluation for a superimposed problem? 3. Is the current treatment working? These questions can only be answered by referring to the natural course of AD, and specifically, information regarding measures of functional impairment and how they change over time. The information that is currently available on this topic is limited and often embodies implicit assumptions that have not been adequately tested. This information will be reviewed, and directions for future research will be outlined. Copyright © 2000 John Wiley & Sons, Ltd.

1. INTRODUCTION

Progressive cognitive impairment is a fundamental characteristic of Alzheimer’s disease (AD). Accurate measurement of progression is essential for the evaluation of therapeutic interventions, and it is also important in many other avenues of research, ranging from studies of clinicopathological relationships to investigations of public health issues. This has prompted the development of a wide variety of assessment instruments, each designed to suit particular research needs. In principle, these instruments could also be extremely valuable in guiding individual patient management, but this application has received relatively little attention. In this review, I will explore the extent to which currently available assessment instruments pertain to the practical management issues that confront patients and their caregivers. I will not evaluate the relative strengths and weaknesses of the instruments as research tools; these have been summarized in several recent reviews [1–5].

* Correspondence to: Douglas J. Gelb, Department of Neurology, University of Michigan Medical School, Ann Arbor, Michigan 48109-0316, U.S.A.
† E-mail: dgelb@umich.edu

Contract/grant sponsor: NIH-NIA; contract/grant numbers: 1-P50 AG08671.

Copyright © 2000 John Wiley & Sons, Ltd.
2. COGNITIVE TESTING

The most direct way to measure severity of dementia is to test cognitive function, either with a comprehensive neuropsychological battery or with one of the abbreviated assessment instruments that have been designed for this purpose [4,6–11]. Results of these tests are often difficult to apply to practical management issues, however. A patient’s ability to copy a drawing of intersecting pentagons does not have obvious implications for caregivers, for example, and the ability to count backwards by seven does not readily translate into a skill that is required in daily activities. Although the practical relevance is more apparent for some of the more fundamental tasks included in the batteries, such as orientation to person, place and time, or the ability to name simple objects, caregivers usually do not require the assistance of a professional to know when patients have deficits in these tasks. At times, fundamental deficits can remain unrecognized because of a patient’s ability to compensate for them, or because of denial or lack of insight on the part of caretakers, but these situations are not common. Conversely, formal cognitive testing can sometimes identify intellectual abilities that are relatively preserved and that can be incorporated by patients and caregivers into compensatory strategies, but it may be just as reasonable to try those strategies empirically and abandon the ones that are not helpful.

Even though most of the items on standard cognitive assessment batteries are not directly relevant to significant management issues, the items might still have clinical value if it could be shown that they function as reliable surrogates for important practical abilities. For example, all patients with AD will eventually reach a point where they do not have adequate visuospatial skills, motor sequencing abilities, reaction times, judgement, or reasoning skills to drive safely, but it can be very difficult to determine precisely when they have reached that point [12–21]. Dementia in and of itself does not preclude the ability to drive. Premature restrictions on driving can undermine a patient’s mobility and sense of autonomy unnecessarily, yet delayed recognition of serious impairment has obvious safety implications both for the patient and for society. Patients often lack the insight necessary to detect their impaired driving performance, and, indeed, they may retain enough ‘instinctive’ ability to perform adequately in routine circumstances. Too often, the first convincing evidence that a patient is too impaired to drive safely comes when the patient is involved in one or more accidents. A reliable indicator of a patient’s capacity to drive safely would be enormously valuable. For the indicator to be clinically useful, a high positive predictive value would be required; a simple statistical correlation would not be sufficient. In fact, even the mini-mental status examination (MMSE) score has a moderate correlation with driving ability, but the relationship is not strong enough to predict driving performance for any individual [14–21]. Measures of reaction time or visual attention may prove to be reliable indicators of driving safety [14,19,20], but this has not yet been established conclusively.

Although driving safety has received the most attention, other important aspects of practical management could conceivably be assessed by using cognitive test instruments [1,22–24]. In principle, there could even be a surrogate measure of legal competence. While that seems unlikely, surrogate measures of ability to cook, handle finances, or other tasks required to live independently would be extremely helpful to the clinician and caregivers, if cut-off values could be established that distinguished competent from incompetent patients with a high sensitivity and specificity. This would require the identification of a measure that had a much stronger correlation with the clinical variable of interest than has been established for any measure to date [25], or else the identification through multivariate analysis of a combination of measures that had high predictive power.
Even if static cognitive measures did not correlate with individual performance of concrete tasks, it is possible that the rate of change of those measures might. Another potential application of the rate of change could be the identification of individual patients with a rate of disease progression that was faster or slower than expected \[1,26–29\]. This could be helpful in evaluating response to a treatment, or in recognizing superimposed disease processes. To be used in this way, a cognitive measure would need to change in a reliable way as AD progresses, with very little variation between patients in the rate of progression \[26,30\].

### 3. GLOBAL ASSESSMENT INSTRUMENTS

A single measure that roughly summarizes a patient’s overall level of impairment can be used to screen candidates for clinical research, to group patients when analysing data, or to assess the outcome of an intervention \[2,4,31–35\]. The value of a global measure in routine clinical practice is less apparent, however. What is to be gained by knowing whether a patient has ‘mild’, ‘moderate’, or ‘severe’ disease? Although caregivers often seem to appreciate having a frame of reference, their response to the information may be as much a function of their own outlook as anything else. Some caregivers will be reassured and comforted by the knowledge that a patient is ‘only stage 1’, for example, whereas others will be depressed by the realization that things can only get worse. Conversely, some caregivers will be disheartened when told that a patient is severely impaired, while others will see it as a validation of the burden they have been experiencing.

There are at least two ways in which global measures could conceivably be useful in clinical practice. First, if there were a treatment strategy that produced a robust improvement in global function but only in a subgroup of patients, then global assessment instruments would provide an efficient way to decide which patients should continue to receive the treatment. Second, if the severity of AD, as assessed by a global assessment instrument, progressed in an extremely reproducible way, then the rating could be used both for prognostic purposes and to detect departures from the expected rate of progression due to superimposed disease processes or a therapeutic intervention.

Before a global scale can be used to direct patient management, however, the assumptions underlying the scale must be considered. By definition, global scales combine information about patients’ performance across a wide range of cognitive domains. The way in which this information is combined necessarily reflects implicit or explicit assumptions about the disease process. As an illustration, consider the Global Deterioration Scale (GDS), a widely used seven-point scale in which each stage is defined by describing representative deficits that a patient at that stage would typically exhibit \[31\]. For example, a patient at stage 5 (moderately severe cognitive decline) ‘may have difficulty counting backward from 40 by 4s or from 20 by 2s’. People at this stage ‘require no assistance with toileting or eating but may have some difficulty choosing the proper clothing to wear...’. They ‘invariably know their own names and generally know their spouses’ and children’s names’. This presupposes that the deficits of AD progress in an orderly fashion in parallel across cognitive domains. If there are any patients who are incontinent or unable to remember their own names but who still retain the ability to calculate, it is not clear how those patients should be classified.

One way to address this problem is to assign a separate rating to a patient’s performance in each of a variety of domains, and then to combine these ratings into a global score. This approach still presupposes that the deficits within a particular domain progress in a predictable manner.
order in patients with AD, but this is more a reflection of the complexity of specific tasks than an assumption about the disease itself. An example of this type of instrument is the Clinical Dementia Rating (CDR) [32], in which patients receive a score of 0 (no impairment), 0.5 (questionable impairment), 1 (mild impairment), 2 (moderate impairment), or 3 (severe impairment) in six independently rated categories: memory; orientation; judgement and problem solving; community affairs; home and hobbies; and personal care. As with the GDS, each score is defined by describing representative deficits, but this time each description is confined to deficits within a single domain. The global CDR is computed from the six component scores using an algorithm in which memory is considered the primary score and all others are secondary.

Unlike the GDS, in which assumptions about the progression of deficits across domains are inherent in the structure of the rating system, the structure of the CDR does not intrinsically require any assumptions about disease progression. The algorithm for computing the global rating does embody such assumptions, however. Indeed, when the CDR was originally published [32], the authors assumed that the secondary scores would always cluster tightly around the memory score. As a result, the algorithm does not permit calculation of an unambiguous global score for many combinations in which the component scores are somewhat dispersed [36].

When the CDR was adopted by the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD), the algorithm was presented in a way that eliminated these indeterminate cases. None the less, the new formulation of the algorithm still incorporated the assumption that the component scores would always cluster tightly around the memory score. If this assumption is invalid, then some counterintuitive and undesirable situations can occur [36]. For example, a decrease in a component score can produce an increase in the global score. Indeed, this type of inconsistency occurs in 5 per cent of all near-match pairs (two sets of component scores that are identical except for a single component score) – the set with the higher component score is assigned a lower global CDR. In an additional 1 per cent of near match pairs, the algorithm produces an anomaly in which the global score increases by two points or more when a component score increases by one point or less. This is not simply a theoretical concern; these undesirable combinations represented 12 per cent of the CDR determinations in one registry of patients with AD, and were also present in the CERAD database [36,37]. A modified version of the CDR algorithm was subsequently published [38], but it failed to eliminate this type of inconsistency [37].

Alternative algorithms exist. The ‘sum of boxes’ score [39], which is conceptually equivalent to calculating the mean of the six category scores, avoids logical inconsistencies but abandons the original emphasis on the memory score. Another published algorithm retains the emphasis on the memory score and eliminates logical inconsistencies by defining the global CDR as the median of the six category scores; when the median falls between two category scores, the one closer to the memory score is chosen as the global CDR [36].

Any attempt to develop a single numerical measure to summarize something as complex and multifaceted as the severity of AD inevitably requires assumptions about disease progression. The assumptions underlying a global assessment instrument may not be immediately apparent, but they should be carefully analysed before the instrument is adopted.

4. FUNCTIONAL ASSESSMENT INSTRUMENTS

Functional assessment instruments are attempts to quantify patients’ abilities to perform routine practical tasks, including self-care activities (ADLs, or activities of daily living) and tasks that
require the use of tools (IADLs, or instrumental activities of daily living) [4, 40–47]. Unlike cognitive testing, for which practical clinical relevance is a significant concern, functional assessment measures translate readily into concrete issues of patient management. The main thing that limits the clinical utility of functional assessment scales is the fact that the information is derived almost exclusively from caregivers’ reports. Although some ADLs and IADLs can be evaluated directly in a controlled setting [44, 45], this is time-consuming and not necessarily representative of patients’ performance in their usual environment. Thus, functional assessment instruments typically require questioning a caregiver. Since a clinician caring for a patient would do this routinely, it is not clear that there is anything to gain by coding the information in terms of the discrete, standardized categories required for a formal assessment instrument. One potential advantage of using the assessment instrument could be that it would ensure that the clinician did not forget to ask about significant practical issues, but this objective could also be met by using a checklist of topics, without constraining the clinician to conform to the response categories of a specific instrument.

A functional assessment instrument could conceivably be useful to a clinician trying to determine whether a patient had improved in response to a medication or some other therapeutic intervention, but it is probably better to use cognitive tests or global scales for this purpose because placebo effects could cast doubt on a measure that is based purely on caregiver reporting. In principle, functional assessment instruments that had predictive power could be useful to the clinician. For example, if there were evidence that a particular pattern on an assessment instrument reliably identified those individuals who would have difficulty with toileting within a year, that would be very useful in counselling caregivers about the need to plan for support services or placement. As with cognitive tests and global scales, a general statistical correlation would not be adequate for this purpose; to be applicable to patient care, the pattern would have to predict with a high degree of confidence when an individual patient would reach a specific endpoint, preferably an endpoint limited to a single functional domain. Some functional assessment scales resemble global rating scales in that they group performance across domains [41, 48]; as with the global scales, the grouping procedure necessarily embodies assumptions about disease progression that could result in misleading predictions.

5. SEVERE IMPAIRMENT BATTERIES

For research purposes, it is important to have measures of disease progression even late in the course, when performance on many clinical, neuropsychological and functional tests tend to show floor effects [49, 50]. By the time these stages are reached, however, it is usually not particularly important for the clinician to have an accurate measure of disease progression. The relevant information can generally be obtained by asking a few simple questions about basic daily functions. Thus, the severe impairment batteries that have been developed have very little clinical utility. The most likely exception would be if one of these batteries were shown to be strongly predictive (on an individual level, not just a group level) for mortality.

6. BEHAVIOURAL RATING SCALES

Behavioural manifestations of AD, such as depression, agitation, hallucinations and delusions, can have profound significance for patient management, but they have typically been ignored in
the kinds of rating scales discussed above. There have been several recent attempts to develop rating instruments for these aspects of the disease [3,51–54]. As with the functional assessment instruments, these rely on caregiver reporting, so in principle they could be redundant for clinicians who are already obtaining complete histories from caregivers. In practice, however, clinicians probably tend to ignore many of these behavioural problems unless caregivers think to report them. Moreover, clinicians may not find these issues as intuitive as they find ADLs and IADLs, so the specific response categories provided by a rating instrument are probably useful. In principle, behavioural rating scales could also be clinically useful if there were evidence that a certain pattern of results reliably identified when individual patients were likely to experience a specific outcome.

7. FUTURE DIRECTIONS

It would be very helpful for clinicians to be able to predict when patients with AD are likely to lose the ability to perform specific functions. The instruments that are already available for measuring disease progression might provide one way to accomplish this, but only if specific measures could be shown to have predictive power at an individual level, not just at a group level. An essential first step would be to determine which aspects of AD are most distressing to patients and caregivers, in terms of both emotional and financial burden [54–56]. Functional assessment instruments that optimally address those functions could then be selected as outcome measures, making each behaviour an independent outcome (rather than grouping functions together and incorporating unnecessary and unwarranted assumptions).

It would also be helpful for clinicians to know when their patients with AD deviated significantly from the expected pattern of disease progression. For this purpose, it is essential to use global instruments that do not impose unwarranted assumptions about disease progression. Here, too, general statements about group averages are of limited clinical utility; the ideal must be to identify individual patients who are progressing in ways that are outside the expected range of variation for the population.

ACKNOWLEDGEMENTS

This work was supported in part by the Michigan Alzheimer’s Disease Research Center, NIH-NIA grant 1-P50 AG08671.

REFERENCES