The evaluation of cognitive function in the dementias: methodological and regulatory considerations

Keith A. Wesnes, PhD; John E. Harrison, PhD

Dementia has as its central feature impairment in cognitive function. Clinically, the cognitive deficit will most often manifest itself as memory problems and most usually as difficulties in the ability to retain new information. However, a number of other areas of cognition are affected and it is important to realize that memory is but one of the cognitive skills compromised in dementia. Dementia is thus prima facie a disorder of cognition and it is our cognitive facilities that underlie our abilities to engage successfully in the activities of daily living (ADL). From this it follows that enhancement to cognitive function will facilitate performance of these ADL. The assessment and understanding of these impairments are crucial to any treatment of the disorder.

Behavioral observation today has a very limited role to play in the assessment of mental ability, as it is now known that many important aspects of cognitive function are not readily assessable by this technique. Cognition can only truly be assessed accurately through the direct use of objective psychometric tests. However, historically the diagnosis of dementia has been largely the province of gerontologists and old-age psychiatrists, who, in the

Impairment of cognitive function is the central feature of dementia. Although, clinically, the cognitive deficit most often manifests itself as memory problems, a number of other areas of cognition are affected, and memory is but one of the cognitive skills compromised in dementia. Dementia with Lewy bodies, for example, accounts for 15% to 25% of all dementias and does not have memory deficits as a core feature. Our cognitive facilities underlie our abilities to engage successfully in the activities of daily living (ADL) and it follows that enhancement of cognitive function will facilitate performance of ADL. The assessment and understanding of these impairments are crucial to any treatment of the disorder. Unfortunately, the principal instrument used to assess cognitive function in most of the major clinical trials of Alzheimer's disease in recent years, the Alzheimer's Disease Assessment Scale–Cognitive Subsection (ADAS-COG), primarily assesses aspects of memory, which has resulted in other important cognitive deficits in dementia being overlooked. Automated cognitive tests are now available that can identify an earlier onset of improvements in dementia in smaller samples than the ADAS. Regulatory authorities should encourage—or even require—the use of automated procedures alongside the ADAS in pivotal trials to help determine the relative utility of the instruments in the fairest way possible. Whatever the outcome, this will be of long-term benefit to patients, carers, drug developers, clinicians, and regulators in this important area.

Dialogues Clin Neurosci. 2003;5:77-88.

Keywords: *cognitive testing; automated testing; cognitive function; dementia; Alzheimer's disease; dementia with Lewy bodies*

Author affiliations: Cognitive Drug Research Ltd, Reading, UK (Keith A. Wesnes); Human Cognitive Neuroscience Unit, Northumbria University, Newcastle upon Tyne, UK (Keith A. Wesnes); Cambridge Psychometric Consultants, Warminster, Wilts, UK (John E. Harrison)

Address for correspondence: Keith A. Wesnes, PhD, Cognitive Drug Research Ltd, Portman Road, Reading RG30 1EA, UK
(e-mail: Keithw@cdr.org.uk)

Selected abbreviations and acronyms

AD	*Alzheimer's disease*
ADAS-COG	*Alzheimer's Disease Assessment Scale—Cognitive Subsection*
ADL	*activities of daily living*
CANTAB	*Cambridge Neuropsychological Test Automated Battery*
CDR	*Cognitive Drug Research*
CNTB	*Computerized Neuropsychological Test Battery*
CPMP	*Committee for Proprietary Medicinal Products*
DLB	*dementia with Lewy bodies*
EBI	*economic buying influence*
EMEA	*European Agency for the Evaluation of Medicinal Products*
EWP	*Efficacy Working Party*
HD	*Huntington's disease*
MCI	*mild cognitive impairment*
MMSE	*Mini-Mental State Examination*
NfG	*note for guidance*
NPI	*Neuropsychiatric Inventory*
SKT	*Syndrome Kurtz Test*
VaD	*vascular dementia*

absence of widely accepted tasks, relied primarily on their clinical judgment. Memory deficits are one of the more obvious aspects of cognitive disorder, both to the patient and the observer; therefore, the marked anterograde amnesia seen in Alzheimer's and other dementias became the hallmark of the disorders. This legacy remains today and can be evidenced by inspecting the *Diagnostic and Statistical Manual of Disorders, Fourth Edition (DSM-IV)* guidelines for any of the dementias.[1] Nonetheless, as will emerge in this chapter, memory deficits are only one of several major aspects of cognitive dysfunction in dementia. Unfortunately, the principal instrument used in most of the major clinical trials of Alzheimer's disease (AD) in recent years, the Alzheimer's Disease Assessment Scale–Cognitive Subsection (ADAS-COG), primarily assesses aspects of memory, which has resulted in other important cognitive deficits in dementia being overlooked. As a consequence, the full therapeutic potential of the treatments studied has not been evaluated.

Assessment of cognitive function

Cognitive functions are those aspects of mental activity that underpin the quality with which we are able to conduct ADL. A number of these aspects of mental activity are subject to change in the efficiency with which they operate. These include attention, short-term (working) memory, long-term memory, reasoning, the coordination of movement, and the planning of tasks. Besides changes due to diurnal rhythms, a wide range of external and internal events can affect the operation of these cognitive functions, including anxiety, fatigue, aging, trauma, disease, psychiatric illness, drugs, hormones, cardiac function, and of course, dementia.

Cognitive function is assessed by requiring subjects (volunteers or patients) to perform specific tasks. The quality of measurement depends on how well the performance of the tasks can be assessed. This is a very important issue, as the process of precisely estimating an individual's level of cognitive competence is affected by a number of sources of error variance. Clearly, when measuring performance, we are hoping to obtain a reliable and precise estimate of an individual's cognitive competence. There are two principal applications of cognitive function tests in clinical practice and research. The first is to identify the ability to conduct the tasks in order to make an assessment of the cognitive capabilities of the particular individual. An obvious example in the context of this article would be to determine the presence and, possibly, degree of dementia. The second is to assess change in cognitive function, ie, to assess a person more than once in order to determine whether the quality of function has altered during the time between the assessments. The latter application is crucial in trials of dementia therapies in which the desire is to determine whether cognitive function has been affected by the therapy.

Individuals vary widely in the quality of their various mental skills and simply assessing them after treatment provides little insight into the nature or extent of any changes. The key to such work is to assess the abilities of the individual prior to treatment and then determine the extent to which these have changed in subsequent assessments. However, repeating cognitive testing in this manner places very stringent constraints on the design and types of tests that can be used. These constraints are not present in many other fields, for example, the repeated assessment of biometric measures such as blood pressure or body weight. With psychometric assessments, which include cognitive tests, performance can change with repeated testing for a variety of reasons that are independent of the study treatment. Examples of these are as follows.

- Learning specific items in memory tests.
- Developing strategies to improve performance.
- Becoming less anxious.
- Improving cognitive skills via training effects.
- Better understanding of the task requirements.

Test developers seek to overcome such effects by developing parallel forms of the tests, for example, having different sets of items to be learned in memory tests, or unpredictable sequences of events in tests of attention. Some forms of tests simply cannot be used, ie, those involving a single strategy, which, once learned, cannot be repeated, or those which, like video games, have no ceiling on practice effects. When sufficient parallel forms are available, evaluations can be conducted of the number of familiarization sessions that are required. Unwarranted test anxiety (some poorly designed tests can be anxiogenic), full understanding of the task requirements, and the determination of optimal strategies can, for many tests in current use, be overcome by two to four repetitions.

Another important control in test design and administration is to ensure that changes in performance of the tasks reflect the quality of the particular aspect of function under study, and not peripheral changes such as alterations to visual function. This can often be achieved by making stimuli large enough that alterations to acuity, for example, will not noticeably affect performance. As there are a variety of independent cognitive functions that need to be assessed in clinical trials, tests should ideally be as specific as possible to particular aspects of cognitive function. Also, it is essential that all aspects of performance that are important in the execution of a task be assessed. For example, if the ability to recognize previously presented items is being measured, the time taken to make the decisions should be precisely recorded. Amazingly, this is done in very few memory tests. This is rather akin to attempting to assess intelligence by requiring volunteers to solve problems, but either not introducing a time constraint or not measuring how long it takes to solve the problems.

Automating cognitive function testing in dementia

The proper automation of cognitive tests is the only way forward in clinical research. Automating cognitive tests can help to overcome many of the problems described in the preceding section, greatly facilitating and standardizing test administration, as well as enhancing test sensitivity. Further, the ability to precisely record cognitive decision times in properly computerized tests can also enable aspects of function to be assessed, such as attention and the speed of memory, which are simply not definitively measurable with pencil and paper tasks.

A wide variety of nonautomated tests have been used in the assessment of dementia. These include the Kew Test, the Kendrick Test, the Mattis Dementia Rating Scale, the Folstein Mini-Mental State Examination (MMSE),[2] the Syndrome Kurtz Test (SKT), and the Alzheimer's Disease Assessment Scale (ADAS). However, computerized tests have also been developed, the three most widely used being the Cambridge Neuropsychological Test Automated Battery (CANTAB),[3] the Computerized Neuropsychological Test Battery (CNTB),[4] and the Cognitive Drug Research (CDR) Computerized Assessment System.[5]

An illustration of the superiority of computerized tests system to nonautomated tests in dementia came from a trial in which Mohr et al[6] contrasted the CDR system with ADAS-COG, the Mattis Dementia Rating Scale, the Wechsler Memory Scale, and the MMSE. The purpose of the study was to contrast the relative utility of the various systems in differentiating patients with Huntington's disease (HD) and AD from each other, as well as from a control group. The authors concluded that the CDR system was able to reliably discriminate the two types of dementia, whereas the other assessments described above were not. Further, in terms of the ability of the various tests to accurately classify the three groups, the computerized tests scored best overall, being able, for example, to accurately identify 77% of the AD patients, compared with the ADAS-COG, which managed to classify 67% correctly. Another notable superiority was 86% accuracy in classifying HD patients with the automated tests, in comparison to 43% with the ADAS-COG, little better than chance. Mohr et al concluded that the assessment of cognitive speed possible with computerization was an important factor in the superiority of the automated system to the other tests.[6]

The International Working Group on Harmonization of Dementia Drug Guidelines has formally recognized the importance of automated cognitive testing in dementia research.[7] In a position paper on "objective psychometric tests in clinical trials of dementia drugs," the group acknowledged the utility of computerized testing:

Automated testing can have clear advantages for clinical trials in this field. The task information is always presented in a standard fashion; the recording of responses is done automatically and precisely, without any bias; and there are no grey areas involving differences of interpretation. These advantages can reduce variability both from session to session for a patient, and also between different national and international sites. Automated procedures have been shown to be more sensitive than the standard tests that are used extensively in this field.

For a detailed comparison of computerized versus pencil and paper assessment testing see Wesnes et al.[8] Another important landmark from the position paper was that it acknowledged that the importance of deficits to attention and information processing in the cognitive symptomatology of AD and other dementias had been largely overlooked, and identified these as domains which should in future be assessed in AD trials. The group also recognized that the ADAS is not appropriate for mildly impaired or at-risk populations.

As speed is such a crucial assessment in cognitive testing, everything possible should be achieved to ensure that it is assessed as accurately as possible. Software should be able to resolve reaction times to the nearest millisecond, which, it should be noted, is not the same as simply giving a score in milliseconds, but with a resolution of say 50 ms. Everything should be done to get the response recorded as quickly as possible, which would, for example, involve avoiding the patient using the keyboard, as the in-built software of PCs only samples this once every 30 ms or so. Also, it should be the patient making the response, not the tester. The use of touchscreens, while potentially of benefit in some types of test, must be carefully managed. The very nature of touchscreens requires the subject to move his/her responding digit to the screen in order to record response time. This task requirement runs the risk of introducing significant levels of error. For example, repeated assessment of this kind can introduce significant fatigue in elderly subjects. A further essential task requirement is to ensure that the starting finger position be consistent both within and between subjects. Some touchscreen-based tests measure reaction time (ie, the time taken to release a home key) and movement time (ie, the time taken to reach a target on the touchscreen). This is a useful decomposition of performance parameters. However, it is essential that the home key accurately records latencies and is of a type and construction that does not selectively disadvantage specific groups of subjects.

Other important methodological issues are to avoid stressful feedback when patients make incorrect responses and to keep the duration of testing to just a few minutes for each test. Systems that can be administered by nonspecialists are advantageous as this facilitates their use in multiple site trials. Tests should ideally measure core domains of human cognitive function discussed earlier, particularly verbal, pictorial, and spatial memory, working as well as episodic secondary memory, various aspects of sustained and focussed attention, and aspects of planning and executive function. Finally, of course, it is necessary besides these considerations of utility to have evidence of the validity, reliability, and sensitivity of the procedures.

If computerized tests are used in clinical trials, all aspects of data capture and processing must of course be sufficiently documented to allow audit to ensure they comply with International Conference on Harmonisation (ICH) good clinical practice (GCP). If the data from testing is to be submitted to the Food and Drug Adminstration (FDA), all systems that are used to capture, process, and analyze the cognitive data must in addition be fully compliant with FDA 21 Code of Federal Regulations (CFR) Part 11 and FDA guidance for computerized systems used in clinical trials. Developing new systems in compliance with 21 CFR Part 11 and making existing systems compliant are both lengthy and often expensive procedures, which sadly preclude most academically developed tests from playing an important role in drug development.

Finally, it must be accepted that cognitive assessment falls within the current domain of psychology, and that researchers not formally trained in psychology should not be in a position to administer and interpret changes from cognitive tests without the close supervision from a suitably qualified psychologist. This is not a protectionist approach from a territorial sense of others encroaching on the discipline of psychology, rather it is in frustration at the widespread superficial application and questionable interpretation of psychological test results by researchers not qualified to conduct or interpret these tests. Journals should require evidence of the involvement of psychologists in research before accepting for publication papers on trials in which cognitive tests have been utilized. Regulatory authorities should mandate the requirement that cognitive data from clinical trials be gathered and interpreted under the supervision of a suit-

ably qualified psychologist. Providers of cognitive tests should ensure that they do not sell the tests to researchers not qualified in psychology or groups without a suitably qualified psychologist. Such restrictions apply to most other psychological instruments, such as personality and aptitude tests, and there is no reason why they should not be applied to the use of tests in drug development.

The profile of cognitive impairment in dementia

The profile of cognitive impairments in dementia has not traditionally included impairments to attention. This is evident in the *DSM-IV* criteria for all of the dementias,[1] where attentional deficits are not even considered as possible symptoms. Further, the scale developed specifically to assess Alzheimer's patients, the ADAS, does not contain an assessment for attention. As suggested earlier, this oversight was probably the result of physicians relying on their clinical judgment, and thus missing less obvious deficits. However, deficits to various aspects of attention in AD have been reported in the literature since 1989,[5,9] and interest in these deficits has now become widespread.[10-12] Importantly, volunteer trials with drugs that stimulate or block the cholinergic system have shown that attention as well as memory can be influenced by the administration of drugs that directly influence the cholinergic system.[13] Further, cholinergic blockade in volunteers with scopolamine mimics the attentional deficits seen in AD.[14,15] This indicates that the cholinergic system plays an important role in controlling various aspects of attentional function. In AD cholinergic deficits lead to attentional impairment, which is therefore central to the cognitive pathology of the disorder. All the preceding evidence would result in the prediction that the anticholinesterases should enhance attention as well as memory in AD. This is precisely what has occurred in trials that have assessed attention; improvements to attention in AD have been seen with the anticholinesterases tacrine,[16,17] velnacrine,[18] and galantamine.[19]

Other major forms of dementia also have impairments to attention as a core feature of the diseases. In the Mohr et al[6] study cited earlier, the HD patients showed greater impairments to attention on the CDR tests, whereas they had smaller deficits to episodic memory. Patients with vascular dementia (VaD) also have greater attentional deficits than AD patients,[20,21] and show smaller impairments to memory.[20] Dementia with Lewy bodies (DLB) completely breaks the traditional mold in dementia by having attention deficits identified as a core feature of the syndrome.[22] These patients also have larger attentional deficits than AD patients while, like HD and VaD patients, they have relatively preserved episodic memory.[21,23] Further, the attentional deficits seen in DLB are not only quantitatively, but also qualitatively different from those seen in AD. For example, the variability in reaction times in a 90-s computerized test of attention, choice reaction time, can discriminate between the two types of dementia with a sensitivity of 81% and a specificity of 92%.[21] Importantly, different profiles of attentional impairment can be seen in all the major types of dementia (VaD, AD, and DLB[21]), probably reflecting the differing etiologies of the conditions. Thus, DLB can also be differentiated from VaD with a sensitivity of 81% and a specificity of 82%, while AD can be differentiated from VaD with a sensitivity of 64% and a specificity of 77%.[21] Attentional impairments have also been seen in first time diagnosed unmedicated Parkinson's disease patients,[24] Parkinson's dementia,[25] and elderly stroke patients actually free from dementia.[26]

Besides marked deficits to attention,[27] demented patients show marked reductions in the speed with which they can recognize previously presented information (words, pictures, faces[5,6,23]). These deficits are also characteristic of mild cognitive impairment (MCI)[28] and add a further dimension to our knowledge of the cognitive deficit profiles in the dementias that have gone undetected by nonautomated assessments like the ADAS. The behavioral impact of delays in time to retrieve information from working and secondary memory is manifest in a variety of behavioral situations. For example, in social situations, in which not only is the patient clearly forgetful, but even when items are retrieved or objects (or people) recognized, the increased time lag makes social interactions more stressed and unsatisfactory. Other examples are patients not remembering the name of someone until after they have passed by in the street, or remembering to do something too late or in the wrong context.

Historical perspectives on cognitive assessment of dementia

Alzheimer's disease

Since the registration of the anticholinesterase, tacrine, for the symptomatic treatment of AD in the late 1980s,

the cognitive outcome measure most frequently used in clinical drug trials for new dementia drugs has been the cognitive subscale of the ADAS (ADAS-COG). However, the ADAS-COG features some well-recognized deficiencies,[29-31] which, as the following examples will illustrate, have been recognized by the International Working Group on Dementia Drug Guidelines[7]:

A generally acknowledged limitation of the ADAS-COG is that it lacks a subset for attention. […] Given the previously noted importance of assessing attention and processing speed in patients with AD, computerized tests can provide optimal procedures for assessing changes in these functions. […] If clear advantages of computerized procedures are demonstrated, such procedures might supersede existing methods.

This situation has led drug developers to seek more sensitive cognitive outcome measures. Regulators, particularly the Efficacy Working Party of the European Medicines Approval Agency, have also opened the possibility of using other, non–ADAS-COG measures. Clinical trials of drugs developed for the amelioration of dementia and especially AD tend to require large numbers of study participants and are typically of quite long duration. Regulators both in Europe and the USA have specified the collection of extensive safety data in support of an application for a marketing license. For example, Leber has specified that a minimum level of safety information is to be based on data for N=1000 study participants collected over a 6-month period.[32] Furthermore, a subset of at least N=300 participants must be further studied for 1 year or more. However, with respect to showing evidence of efficacy, a combination of modest degrees of drug efficacy and the use of relatively insensitive instruments has meant that typically hundreds of study participants are required for trials lasting at least 6 months and often considerably longer. Added to this situation is the practical and ethical difficulty of recruiting patients for the placebo arm of these trials. These demanding requirements have made large, multicenter, international trials a necessity.

The routine inclusion of the notoriously unreliable clinicians' impression scales is seen as tacit acceptance of the failure of current cognitive outcome measures to capture the clinically significant improvements seen in patients. It therefore seems clear that pretenders to the ADAS-COG's crown will benefit from being demonstrably robust proxy measures of everyday cognitive improvement. Intuitively, it seems reasonable to suppose that enhancements in cognition seen in laboratory-based assessments will be reflected as improvements in day-to-day activities reliant upon reasonable degrees of cognitive competence. One method for validating laboratory-based methods would be to correlate them against concurrently run ADL and quality of life questionnaires. The result of such a validation project may well yield cognitive outcome measures that are powerful and accurate proxy measures of clinically significant drug enhancements. This validation has the potential to make clinicians' rating scales redundant as a means of capturing the positive effects of pharmaceutical interventions.

Dementia with Lewy bodies

DLB accounts for 15% to 25% of all dementia.[33] As described earlier, DLB is a newly diagnosed form of dementia for which consensus criteria have emerged in recent years. Here are the consensus criteria for a diagnosis of probable DLB[34]:

- The central feature required for a diagnosis of DLB is progressive cognitive decline of sufficient magnitude to interfere with normal social or occupational function. Prominent of persistent memory impairment may not necessarily occur in the early stages, but is usually evident with progression. Deficits on tests of attention and frontosubcortical skills and visuospatial ability may be especially prominent.

Plus two of the following essential features:

- Fluctuating cognition with pronounced variations in attention and arousal.
- Recurrent visual hallucinations, typically well-formed and detailed.
- Spontaneous motor features of Parkinsonism.

It is evident from the criteria that attentional deficits assessed by tests are a key part of the diagnosis, and that memory deficits may not be apparent in the early stages. This makes it clear that the ADAS-COG would not be a suitable primary outcome measure in a therapeutic trial of DLB as it does not assess attention or frontosubcortical skills and visuospatial ability. Thus, in the first randomized, placebo-controlled, double-blind trial of an anticholinesterase in DLB,[33] the two primary outcome measures were a compound speed score derived from the CDR cognitive assessment system and a DLB-typi-

cal summary score from the Neuropsychiatric Inventory (NPI). This was a prospective, multicenter, randomized, double-blind, 20-week, placebo-controlled study conducted in the UK, Spain, and Italy. Testing was conducted prior to dosing, at weeks 12 and 20 after the commencement of dosing, and again 3 weeks after dosing was stopped. Analysis of the data from the 92 patients who completed the study identified a significant pattern of benefits of rivastigmine over placebo on the two main outcome criteria. Benefits on the CDR system measures were seen on tests of attention, working memory, and episodic secondary memory. For example, on the speed score from the computerized tests, patients given placebo showed significant deterioration from predosing at weeks 12 and 20, whereas patients on rivastigmine performed significantly above their predosing levels. Some quotes from the parts of a paper[33] prepared by the physicians indicate the growing acceptance of the clinical relevance of measuring the speed of cognitive function:

Since behavioral slowing and severely impaired attentional function are key features of Lewy-body dementia, we used the sum of the latencies for the computerized cognitive assessment tests (speed score) as the second primary efficacy measure. […] Improvements in psychiatric and behavioral features were mirrored by changes in cognitive performance. […] The clinical relevance of these improvements in attention was captured in caregiver reports of patients, describing them as more alert and switched on, and emphasized by reduced apathy scores on NPI.

A commentary in the *Lancet*[35] on this trial also revealed the widening acceptance of computerized tests in dementia research:

The use of reaction times as a second primary outcome measure is another novel feature of this trial. […] Neuropsychological functions other than those evaluated with the ADAS-COG […] are also relevant to the treatment of patients with dementia. […] McKeith and coworkers show that other features, such as psychological symptoms and reaction times, can be meaningful outcome measures in dementia drug trials.

These effects seen in this trial were also large in magnitude: at week 12 a factor score, power of attention, declined by 19% on placebo compared with an improvement of 23% on rivastigmine.[36] From the above, it seems clear that there is little relevance for the ADAS for DLB, except possibly as a secondary measure to compare findings to previous trials with AD. Attention is a core feature of the disease, as is behavioral and mental slowing, which means that assessing attention, speed of access to memory, as well as overall memory performance with a computerized system is clearly optimal.
Another contribution to the estimation of clinical relevance in this trial was that the system used has a large normative database. This has allowed the clinical relevance of these data to be assessed. In this trial, rivastigmine reduced the DLB deficit on the power of attention factor (the difference between the DLB patients and age-matched controls) by 33%.[36] In other words, the attentional impairments in the patients were pushed one third of the way back towards being normal, a large effect size and one for which the clinical relevance is clearly apparent. This should be contrasted with the ADAS, which does not have a database of scores for normals. The only way of assessing the clinical relevance of effects on ADAS-COG is to use the number of points moved in order to estimate how long treatment may prevent the patient from becoming institutionalized. This is obviously important, and the computerized system also has similar longitudinal data and can thus make this assessment; but describing treatment response in terms of the degree to which the patient has been "normalized" is an extremely valuable extra piece of information that has far more intuitive appeal.
This trial confirmed that computerized cognitive tests can be suitable and effective as primary outcome variables in dementia trials. More importantly for DLB, it illustrated that automated tests that incorporate sensitive measures of attention and other cognitive skills not assessed by the ADAS are more suitable primary outcome measures. It is clear from this important trial that for DLB, the ADAS does not have a role as a primary outcome variable in pivotal trials, though it should be included as a secondary measure to enable comparisons to be made to the effects of other treatments in AD.

Regulatory requirements for dementia drug trials

The Efficacy Working Party (EWP) of the European Agency for the Evaluation of Medicinal Products (EMEA) has provided reasonably specific guidelines for the use of cognitive testing in clinical drug trials of compounds for use in dementia.

The "Note for Guidance" (NfG) document published by the EWP states that "improvement of symptoms should be assessed in the following three domains":

- Cognition.
- ADL.
- Overall clinical response.

Little guidance is given with respect to the specific cognitive tests that should be administered and the authors of the NfG acknowledge that:

Whilst a large number of methods for evaluation of cognitive functions and behavioral changes have been suggested, none has convincingly emerged as the reference technique. […] Hence the choice of assessment tools should remain open, provided that the rationale for their use is presented, and justified.

This statement provides for the possibility of using cognitive outcome variables other than the ADAS-COG. Thus, it is possible to consider adopting cognitive tests that have the propensity to show efficacy in fewer patients and in trials that are briefer than the typical ADAS-COG trial. Such an opportunity would be welcomed in early phase 2 trials, where proof of principle and/or optimal dose ranges are sought. Patient numbers in the previously mentioned trials with the CDR system were modest (tacrine, n=32; velnacrine, n=35; galantamine, n=30). The DLB trial mentioned in the previous section involved 92 patients. In a further bridging trial with S12024 in AD, significant cognitive effects with computerized tests were seen in 53 AD patients.[37] Such tests thus have much utility in phase 2 trials, and it is possible to use them even earlier in the development process. In one trial, acute effects of a potential antidementia compound were seen by administering computerized tests prior to dosing and 15, 40, and 240 min afterwards in 12 Alzheimer's patients.[38] The latter trial shows that demented patients can be tested in phase 1 conditions, and opens the possibility for cognitive bridging trials between phases 1 and 2.

It might also be possible to persuade European regulators to grant marketing approval on the basis of results obtained using non–ADAS-COG outcome measures. Clearly, this course of action would benefit from discussion with both the cognitive test provider and the regulators themselves. Experience suggests that a relatively quick and accessibly priced method of soliciting a regulatory opinion is to approach a national agency, such as the UK's Medicines Control Agency, which has proven helpful during recent enquiries.

Further details on cognitive testing requirements for dementia drug trials are given in Section 2.2.1 of the EWP NfG under the heading "Objective cognitive tests":

Objective tests of cognitive function must be included in the psychometric assessment; such tests or batteries of tests must cover more than just memory, as impairments in domains other than memory are mandatory for the diagnosis of AD and the assessment of its severity. Within the domain of memory, several aspects should be assessed. These are learning of new material, remote as well as recent memory, and recall and recognition memory for various modalities (including verbal and visuospatial). Other cognitive domains such as language, constructional ability, attention/concentration and psychomotor speed should be assessed as well.

Thus, the NfG provides for the assessment of a significant number of different cognitive domains, including domains not tested by the ADAS-COG, such as psychomotor speed and attention. This reflects the previous recommendations of the International Working Group on Dementia Drug Guidelines cited earlier.[7] However, no specific guidance is given regarding which particular tests should be used in the cognitive assessment. Instead the authors state that:

The Alzheimer's Disease Assessment Scale (ADAS) cognitive subscale, dealing with memory, language, construction and praxis, orientation, is widely used. However, this remains an open research field.

This appears to underline the EWP's willingness to consider tests other than the ADAS-COG.

Efficacy measurement for trials conducted in the USA

The ADAS-COG has become the "gold" standard for dementia drug trials in the USA, in spite of its acknowledged deficiencies.[6] An attempt has been made to remedy the absence of tests of attention from the original version by the inclusion of two additional nonautomated tests, bringing the total number of subtests to 13. Given the status of the ADAS-COG and its continued apparent popularity, the inclusion of this assessment in pivotal

phase 3 trials of dementia drugs is highly recommended. It should also be included in larger phase 2 trials, though not necessarily as the primary outcome. Here, other more sensitive procedures or tests that cover major domains of function not covered by the ADAS-COG could be considered as primary outcomes, as the purpose of phase 2 trials is to identify optimal doses and dosing strategies, and also of course proof of concept.
In spite of a perception that ADAS-COG is the only acceptable outcome measure for use in AD clinical drug trials, an influential guidance paper published by Leber during his time with the FDA did not mandate the use of the ADAS-COG.[32]

The requirement for coprimary efficacy

Given that dementia is prima facie a disorder of cognition, it at first seems entirely reasonable to consider granting marketing approval to drugs that occasion cognitive improvement. However, an important consideration for regulators is the clinical relevance of the observed cognitive changes. Traditionally, a four-point ADAS-COG advantage of drug over placebo has been seen as sufficient evidence of efficacy for regulators to issue marketing approval. However, recent reviews of the efficacy of licensed drugs have cast considerable doubt on the validity of this assumption. For example, in their 2001 review of dementia drugs, the UK's National Institute for Clinical Excellence[39] stated that:

It is not clear the extent to which cognitive measures such as ADAS-COG or MMSE are accompanied by real-life functional changes that are meaningful to patients and their carers.

This perspective begins to explain why regulators require evidence of positive drug effects on either clinician-rated impression of change scales or ADL scales. Intuitively, it seems reasonable to suppose that enhancements in cognition are likely to be accompanied by improvements in day-to-day functioning. However, data in support of this proposition are sparse and the concern remains that cognitive changes reported using scales such as the ADAS-COG may not be accompanied by clinically relevant functional improvements. Clear evidence that cognitive enhancement reliably accompanied functional improvement might allow us to reduce the role the clinician's rating and/or ADL scale assessments. Evidence from one computerized system is available in a large trial with data available for 744 AD patients. Here the Instrumental Activities of Daily Living scale was administered predosing, together with the computerized cognitive tests. There were highly significant correlations between the ADL scale and the three major factor scores from the computerized system (r=0.43 for power of attention, r=0.39 for speed of memory, and r=0.48 for quality of memory; all P<0.0001). These correlations, while not large in magnitude, clearly identify a direct relationship between these cognitive assessments and how well the patients were judged to cope with everyday activities. In previous work with the same system, correlations of up to 0.79 were seen on individual task measures and the Stockton Geriatric Rating Scale, a scale completed by ward staff concerning the abilities of institutionalized geriatric patients to conduct ADL.[5] As more data of this kind accumulate so will the acceptance grow that changes in tests of cognitive function have clinical significance for everyday behavior.

Overall conclusions and recommendations

- The traditional dementias, AD and VaD, must be acknowledged to be far more than simply disorders of memory. Trials that evaluate the effectiveness of potential therapies need additionally to include sensitive assessments of the other aspects of dysfunction, such as attention.
- DLB accounts for between 15% and 25% of all dementias, and does not have memory deficits as a core feature of the disease. Trials to assess the efficacy of novel treatments for DLB should therefore use cognitive test systems that address the major impairments of disorder, and attentional assessments are particularly relevant here.
- Cognitive tests should only be administered under the direct supervision of individuals suitably trained in psychology, and proof of such supervision should be a regulatory requirement.
- Automated cognitive tests are available and can identify an earlier onset of improvements in dementia in smaller sample sizes than the ADAS. Such tests should thus be used in bridging trials between phases 1 and 2, and also in phase 1 trials to enable smaller and shorter trials to be conducted for proof of concept to be identified; as well as for optimal doses and dosing regimen to be established. However, automated tests will need to satisfy the stringent ICH GCP and FDA require-

ments before they can be used in such work.

- More work needs to be conducted to establish the everyday relevance of tests of cognitive function. Once this is established, the dependence on insensitive daily living and functional ability scales will be reduced and the outcomes in clinical trials will be more appropriate.
- The ADAS-COG is the current gold standard for pivotal trials in AD. This situation leads to a number of major difficulties due to the widely acknowledged inadequacies of the scale. The situation is not dissimilar to that of depression, where the Hamilton Depression Scale has become the "regulatory gold standard" despite its widely recognized numerous shortcomings.

The development of antidementia drugs is, however, in its infancy and there is still time to prevent this field ending up in the same unsatisfactory situation as exists in antidepressant development. To achieve this, regulatory authorities must encourage or even require the use of other automated procedures alongside the ADAS in pivotal trials. This will help determine the relative utility of the instruments in the fairest way possible. Either such work will confirm the ADAS as the optimal tool in the field, or other more suitable tools will be identified. Either outcome will be to the long-term benefit of patients, carers, drug developers, clinicians, and regulators in this important area. ❑

REFERENCES

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: American Psychiatric Association; 1994.
2. Folstein MF, Folstein SE, McHugh PR. "Mini-Mental State": a practical method for grading the cognitive state of patients for the clinician. *J Psychiatry Res*. 1975;12:189-198.
3. Swainson R, Hodges JR, Galton CJ, et al. Early detection and differential diagnosis of Alzheimer's disease and depression with neuropsychological tasks. *Dement Geriatr Cognitive Disord*. 2001;12:265-280.
4. Veroff AE, Cutler NR, Sramek JJ, Prior PL, Mickelson W, Hartman JK. A new assessment tool for neuropsychological research: the computerised neuropsychological test battery. *J Geriatr Psychiatry Neurol*. 1991;4:211-217.
5. Simpson PM, Surmon DJ, Wesnes KA, Wilcock GR. The Cognitive Drug Research computerised assessment system for demented patients: a validation study. *Int J Geriatr Psychiatry*. 1991;6:95-102.
6. Mohr E, Walker D, Randolph C, Sampson M, Mendis T. The utility of clinical trial batteries in the measurement of Alzheimer's and Huntington's dementia. *Int Psychogeriatr*. 1996;3:397-411.
7. Ferris S, Lucca U, Mohs R, et al. Objective psychometric tests in clinical trials of dementia drugs. Position paper from the International Working Group on Harmonization of Dementia Drug Guidelines. *Alzheimer Dis Assoc Disord*. 1997;11(suppl 3):34-38.
8. Wesnes KA, Hildebrand K, Mohr E. Computerised cognitive assessment. In: Wilcock GW, Bucks RS, Rocked K, eds. *Diagnosis and Management of Dementia: A Manual for Memory Disorders Teams*. Oxford, UK: Oxford University Press; 1999:124-136.
9. Nebes RD, Brady CB. Focussed and divided attention in Alzheimer's disease. *Cortex*. 1989;26:305-315.
10. Baddeley AD, Baddeley HA, Bucks RS, Wilcock GK. Attentional control in Alzheimer's disease. *Brain*. 2001;124:1492-1508.
11. Foster JK. Selective attention in Alzheimer's disease. *Front Biosci*. 2001;6:135-153.
12. Perry RJ, Hodges JR. Attention and executive deficits in Alzheimer's disease. *Brain*. 1999;122:383-404.
13. Wesnes K, Warburton DM. Effects of scopolamine and nicotine on human rapid information processing performance. *Psychopharmacology*. 1984;82:147-150.
14. Wesnes KA. The extent of the involvement of cholinergic deterioration in the profile of cognitive impairment seen in Alzheimer's disease. *Proc Br Psychol Soc*. 1996;4:93.
15. Wesnes KA. The pathology of attention of the dementias. *J Psychopharmacol*. 1996;10(suppl):A51.
16. Sahakian BJ, Owen AM, Morant NJ, et al. Further analysis of the cognitive effects of tetrahydroaminoacridine (THA) in Alzheimer's disease: assessment of attentional and mnemonic function using CANTAB. *Psychopharmacology*. 1993;110:395-401.
17. Wesnes K, Scott M, Boyle M, Surmon DJ, Wilcock GK. Use of the Cognitive Drug Research computerised assessment system to measure the efficacy of THA and galanthamine in Alzheimer's disease. *Psychopharmacol Bull*. 1994;30:139.
18. Siegfried KR. Pharmacodynamic and early clinical studies with velnacrine. *Acta Neurol Scand Suppl*. 1993;149:26-28.
19. Wesnes KA, Scott M, Morrison S, Greenwood D, Russell-Duff K, Wilcock GK. The effects of galanthamine on attention in Alzheimer's disease. *J Psychopharmacol*. 1998;12:(suppl A):A46.
20. Ayre G, Ballard C, Pincock C, McKeith I, Sahgal A, Wesnes KA. Double dissociation between dementia with Lewy bodies and Alzheimer's disease on tests of attentional and mnemonic function: the role of the basal forebrain. *J Psychopharmacol*. 1998;12(suppl A):A41.
21. Walker MP, Ayre GA, Cummings JL, et al. Quantifying fluctuation in dementia with Lewy bodies, Alzheimer's disease and vascular dementia. *Neurology*. 2000;54:1616-1625.
22. McKeith IG, Ayre GA. Consensus criteria for the clinical diagnosis of dementia with Lewy bodies. In: Iqbal K, Winblad B, Nishimura T, Takeda M, Wisnieswski HM eds. *Alzheimer's disease: Biology, Diagnosis and Therapeutics*. Chichester, UK: John Wiley & Sons; 1997:167-178.
23. Ayre GA, McKeith IG, Sahgal A, Ballard CG, Wesnes KA. Dementia with Lewy bodies, Alzheimer's disease and vascular dementia show distinct patterns of cognitive impairment. *J Psychopharmacol*. 1997;11(suppl):A56.
24. Moon G, Wesnes KA, Manktelow TC. Cognitive deficits in recently diagnosed untreated patients with Parkinson's disease. *J Psychopharmacol*. 2002;16(suppl):A31.
25. Ballard C, Manktelow TC, Arsland D, Wesnes KA, Starrfelt R. Attention and fluctuating attention in Parkinson's disease with and without dementia and dementia with Lewy bodies. *J Psychopharmacol*. 2002;16(suppl):A31.
26. Ballard C, Wesnes KA, O'Brien J, et al. Cognitive impairments in elderly stroke patients without dementia: profile and MRI correlates. *J Psychopharmacol*. 2002;16(suppl):A31.
27. Wesnes KA, Bullock R. The emerging importance of attention and the speed of cognitive function in ageing and dementia research. *CPD Bull Old Age Psychiatry*. 2001;3:11-15.
28. Nicholl CG, Lynch S, Kelly CA, et al. The Cognitive Drug Research computerised assessment system in the evaluation of early dementia - is speed of the essence? *Int J Geriatr Psychiatry*. 1995;10:199-206.
29. Harrison JE. Cognitive testing and drug development. *Clin Res Focus*. 2001;12:5-11.

La evaluación de la función cognitiva en las demencias: consideraciones metodológicas y para las agencias reguladoras

El deterioro de la función cognitiva es la característica central de la demencia. Aunque clínicamente el déficit cognitivo se manifiesta, la mayoría de las veces, como problema de memoria, hay un número de otras áreas de la cognición afectadas y la memoria es sólo una de las herramientas cognitivas comprometidas en la demencia. La demencia con cuerpos de Lewy, por ejemplo, da cuenta del 15% a 25% de todas las demencias y no tiene déficits de memoria como característica central. Nuestras destrezas cognitivas subyacen a nuestras capacidades para realizar con éxito las actividades de la vida diaria (AVD) y al continuar con la estimulación de la función cognitiva se facilitará la ejecución de las AVD. La evaluación y comprensión de estos deterioros son cruciales para cualquier tratamiento de este trastorno. Lamentablemente, el principal instrumento utilizado para evaluar la función cognitiva en la mayoría de los ensayos clínicos más importantes en la Enfermedad de Alzheimer en los años recientes, la subsección cognitiva de la Escala de Evaluación de la Enfermedad de Alzheimer (ADAS-Cog), evalúa primariamente aspectos de memoria, lo que ha significado que se hayan descuidado otros déficits cognitivos en la demencia. Actualmente se dispone de pruebas cognitivas automatizadas, las cuales pueden identificar un inicio más precoz del deterioro en la demencia, en muestras más pequeñas que para la ADAS. Las autoridades reguladoras deben favorecer- o aun requerir- el empleo de procedimientos automatizados junto con la ADAS en los ensayos fundamentales, para ayudar a determinar la utilidad relativa de los instrumentos de la forma más adecuada posible. Cualquiera que sea la evolución en esta área importante, se traducirá en un beneficio a largo plazo para los pacientes, los cuidadores, los investigadores de fármacos, los clínicos y las agencias reguladoras.

Évaluation de la fonction cognitive dans les démences : considérations méthodologiques et régulatrices

L'altération de la fonction cognitive est l'élément central de la démence. Bien que, cliniquement, le déficit cognitif se manifeste le plus souvent par des problèmes de mémoire, un certain nombre d'autres domaines de la cognition sont atteints et la mémoire n'est qu'une des aptitudes lésées lors de la démence. La démence avec corps de Lewy, par exemple, compte pour 15 % à 25 % de toutes les démences et les déficits de la mémoire n'en sont pas la caractéristique principale. Nos structures cognitives sous-tendent nos capacités à réaliser avec succès les tâches quotidiennes et il s'ensuit que la stimulation de la fonction cognitive facilitera l'exécution de ces tâches. L'évaluation et la compréhension de ces détériorations sont fondamentales quel que soit le traitement du trouble. Malheureusement, l'outil principal utilisé pendant les dernières années pour évaluer la fonction cognitive dans la plupart des essais cliniques de la maladie d'Alzheimer, l'ADAS-COG (Alzheimer's Disease Assessment Scale-Cognitive Subsection)*, évalue surtout les aspects de la mémoire, ce qui a conduit à négliger d'autres déficits cognitifs importants de la démence. Des tests automatisés cognitifs sont maintenant disponibles et ils peuvent déceler un début plus précoce des améliorations lors de la démence avec de plus petits échantillons que pour l'ADAS. Les autorités réglementaires devraient encourager, ou même exiger, l'utilisation de procédures automatisées à côté de l'ADAS dans des essais pivots pour permettre de déterminer l'utilité relative des outils de la façon la plus juste. Quels que soient les résultats, cela apportera un bénéfice à long terme pour les patients, le personnel soignant, l'industrie pharmaceutique, les cliniciens et les instances réglementaires dans cet important domaine.*

30. Harrison JE. Routine cognitive testing for all drugs? *Drug Discovery Today*. 2002;7:101-102.

31. Veroff AE, Bodick NC, Offen WW, Sramek JJ, Cutler NR. Efficacy of xanomeline in Alzheimer disease: cognitive improvement measured using the Computerized Neuropsychological Test Battery (CNTB). *Alzheimer Dis Assoc Disord*. 1998;12:304-312.

32. Leber P. *Guidelines for the Clinical Evaluation of Antidementia Drugs*. Bethesda, Md: Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation Research; 1990.

33. McKeith IG, Del Ser T, Spano F, et al. Efficacy of rivastigmine in dementia with Lewy bodies: results of a randomised placebo-controlled international study. *Lancet*. 2000;356:2031-2036.

34. McKeith IG, Galasko D, Kosaka K, et al. Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): Report of the consortium on DLB international workshop. *Neurology*. 1996;47:1113-1124.
35. Cummings JL. Cholinesterase inhibitors: expanding applications. *Lancet*. 2000;356:2024-2025.
36. Wesnes K, McKeith IG, Ferrara R, et al. Effects of rivastigmine on cognitive function in dementia with Lewy bodies: a randomised placebo-controlled international study using the Cognitive Drug Research computerised assessment system. *Dement Geriatr Cognitive Disord*. 2002;13:183-192.
37. Allain H, Neuman E, Malbezin M, et al. Bridging study of S12024 in 53 in-patients with Alzheimer's disease. *J Am Geriatr Soc.* 1997;45:125-126.
38. Templeton L, Barker A, Wesnes K, Wilkinson D. A double-blind, placebo-controlled trial of intravenous flumazenil in Alzheimer's disease. *Hum Psychopharmacol*. 1999;14:239-245.
39. NICE. Technology Guidance Appraisal No. 19: Guidance on the use of donepezil, rivastigmine and galantamine for the treatment of Alzheimer's disease. London, UK: National Institute for Clinical Excellence; 2001.