The role of human cognitive neuroscience in drug discovery for the dementias
Keith A Wesnes1,2,3 and Chris J Edgar4

Cognitive dysfunction characterizes all the various forms of dementia. Evidence is accumulating that all of the progressive neurodegenerative dementias, such as Alzheimer’s disease (AD), are preceded by years, if not decades, of pathological cognitive decline. The limited effectiveness of the four current medications registered for AD together with the failure of dozens of programmes over the last decade has influenced the decision to evaluate treatment at earlier stages of the disease; even before any cognitive symptoms have appeared. However, it has to be acknowledged that treating mild cognitive impairment (MCI) as a prodrome for AD has also had very limited success. Nonetheless a more important problem in MCI research, and dementia in general, has to be laid at the door of the limited effectiveness of the cognitive tests employed. This problem will become even more severe for the latest research direction of treating preclinical AD because such individuals will have levels of cognitive abilities which are in the normal range; and thus many of the scales currently used in dementia research will not be sufficiently demanding to identify change over time. This paper reviews and discusses the methodology and instruments available for research and clinical practice in this major area; with a focus on the challenges involved in test selection and evaluation.

Addresses
1 Bracket, Goring-on-Thames, UK
2 Department of Psychology, Northumbria University, Newcastle, UK
3 Centre for Human Psychopharmacology, Swinburne University, Melbourne, Australia
4 Formerly Bracket, Goring-on-Thames, UK

Corresponding author: Wesnes, Keith A
(Keith.wesnes@bracketglobal.com)

Introductions
The search for substances to improve cognitive function and treat either normal or pathological age-related cognitive decline is entirely dependent upon the availability of instruments which can reliably measure such effects. Alzheimer’s disease (AD) is the most common form of neurodegenerative dementia, and some success in temporarily slowing the rate of cognitive decline with the four registered treatments for the disease has been identified [1**,2]. However, over the last 10 years the field has been beset by dramatic failures of numerous very large and hugely expensive intervention trials with a range of novel treatments including potential vaccines [3**,4*]. Part of the blame for these failures has been levelled at the failure of the major clinical scales used to assess cognition in the trials to properly reflect the cognitive status of the patients (e.g. [4*]). This limited progress has resulted in the major focus of the field shifting from waiting for the symptoms of cognitive dysfunction to become fully manifest before initiating treatment—to the evaluation of novel therapeutic interventions at much earlier stages of the disease. A crude analogy could be no longer waiting for flames to emerge from upper windows of a building before acting; but instead attempting to extinguish the small fire in the basement which preceded it. Thus the focus of therapeutic trials in AD has now broadened to encompass the prodromal and even preclinical stages of the disorder, that is, moving from treating symptoms to prophylaxis. This move was initially stimulated in 2007 by diagnostic criteria designed to capture earlier stages on the basis of early episodic memory loss alongside biomarker evidence of disease pathology [5], and these criteria have been further refined [6]. This has been consolidated by recent publications from workgroups convened by the National Institute on Ageing (NIA) and the Alzheimer’s Association, one which provided guidelines for the diagnosis of mild cognitive impairment (MCI) because of AD [7**], and another which proposed a conceptual framework and operational research criteria for the preclinical stage of AD [8**].

The revised criteria now recognize MCI because of AD, which the group termed the symptomatic predementia phase of AD [7**], also referred to as prodromal AD. This recognition that AD may be diagnosed at stages before dementia, and the requirement to properly assess the resulting impairment to cognition and everyday function, has placed far greater emphasis on cognitive/psychological assessment and has made the selection and interpretation of such assessments both more critical and challenging. The guidance makes a number of recommendations regarding cognitive assessment, and whilst mentioning particular tests of certain cognitive domains, falls short of any recommendations of cognitive test selection and their specific implementation, for example,
use of particular cut-offs in diagnosis [7**]. On the positive side, the field now has guidance regarding the range of possible cognitive symptoms that may be associated with AD dementia as a third workgroup revised the 1984 NINCDS-ADRDA criteria for the diagnosis of dementia because of AD [9*]. This new guidance represents a shift from memory impairment as being the primary impairment in the condition, allowing non-amnestic presentations of the disease to also be targeted; including deficits to attention, information processing, visuospatial skills and executive function [9*].

Preclinical AD has been defined as the progression from normal cognition to MCI and AD, acknowledging that subtle cognitive changes in a number of domains may precede these stages [8**]. However, the preclinical guidelines were intended for research purposes with no clinical/diagnostic utility ascribed them. An important paradigm shift was the recognition that early diagnosis measured by change in cognition over time would be a far more appropriate approach than one-time assessment; and that longitudinal studies of older individuals, using measures sensitive to detecting very subtle cognitive decline and/or improvement would be required [8**]. Despite the preclinical guidelines claiming to have no clinical/diagnostic utility, research has nonetheless progressed. For example a five year 5000+ participant industry sponsored preclinical AD therapeutic trial has begun; the study being designed to prevent the occurrence of MCI [10]. Three other programmes are planned or underway (4; see Table 1) The A4 (anti-amyloid treatment in asymptomatic Alzheimer’s disease prevention) trial has aroused much interest and will start in late 2013 [3**]. This secondary prevention trial in 1000 clinically normal older individuals (aged 70–85) who are positive on biomarkers associated with the risk of developing AD will be treated for three years. The trial is funded by the National Institutes of Health (NIH), part of the US Department of Health and Human Services, and by private sector contributions. This year it has been announced that the A4 trial has selected Solanezumab (a monoclonal antibody compound) as the treatment (http://www.fiercebiotech.com/press-releases/researchers-announce-treatment-choice-alzheimers-disease-a4-prevention-clin).

Regulatory guidance is essential in this as in all CNS fields, and the US Food and Drug Administration (FDA) has been proactive in this area. In February 2013 the Division of Neurology Products of the Agency issued comments a draft Guideline for Industry entitled ‘Alzheimer’s Disease: Developing Drugs for the Treatment of Early Stage Disease’ (www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM338287.pdf). The document outlined the FDA’s current thinking regarding the selection of patients with early AD, or patients who are determined to be at risk of developing AD, for enrolment into clinical trials. The guidance also addressed the selection of endpoints for clinical trials in these populations, as well as the manner in which disease modification might be demonstrated. This was followed up the next month, by a New England Journal of Medicine article, authored by FDA officers, which made the important point that at the earliest stages of the disease (i.e. preclinical AD), functional impairment would be difficult to assess; proposing that it could be feasible to approve a drug on the basis of assessment of cognitive outcome alone, using the FDA’s accelerated approval pathway [11**] (see Figure 1). This is a major breakthrough both for researchers and pharmaceutical companies who would prefer to employ sensitive cognitive tests in this area, and has aroused huge interest and debate.

This review will summarize some of the key considerations in the selection and application of cognitive tests for use in the diagnosis of neurodegenerative disorders before the stage of dementia. In particular, there will be a focus on the relative benefits of longitudinal assessment to identify a decline in function over time in the individual, versus ‘one-off’ assessment and comparison to normative data. Detailed consideration will also be given to the properties of cognitive tests which could serve as effective outcome measures in both research and therapeutic trials in preclinical dementia as well as MCI.

### Table 1

<table>
<thead>
<tr>
<th>Trial</th>
<th>Main patient population</th>
<th>Number of subjects*</th>
<th>Drugs (all versus placebo)</th>
<th>Primary aim and trial duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>API</td>
<td>Asymptomatic PSEN1 E280A carriers, within 15 years of expected age of onset</td>
<td>216</td>
<td>Crenezumab</td>
<td>Cognition at five years</td>
</tr>
<tr>
<td>DIAN</td>
<td>Asymptomatic ADAD mutation carriers, 15 years before and up to 10 years after expected age of onset</td>
<td>160</td>
<td>Three undisclosed drugs</td>
<td>Target engagement at two years; a subsequent trial will assess cognition three years on Cognition at three years</td>
</tr>
<tr>
<td>A4</td>
<td>Asymptomatic elderly patients with PET-amyloid positivity</td>
<td>1000</td>
<td>An undisclosed drug</td>
<td></td>
</tr>
</tbody>
</table>

(Taken from [3**])

* Only including mutation carriers and positron emission tomography (PET)-amyloid positive-patients; all trials will also enrol negative biomarker patients for ethical reasons. A4, anti-amyloid treatment in asymptomatic Alzheimer’s disease; ADAD, autosomal-dominant Alzheimer’s disease; API, Alzheimer’s Prevention Initiative; DIAN, Dominantly Inherited Alzheimer Network; PSEN1, presenilin 1.
**The identification of abnormal cognitive decline**

The next sections will consider the relative merits of identifying preclinical dementia in a ‘one-off’ cross-sectional manner versus a longitudinal manner.

**One-off/cross-sectional assessment**

Long term studies following subjects with MCI have demonstrated an increased likelihood of progression to dementia (of all types) versus age matched normal controls. In such studies following application of other criteria besides impairment on cognitive tasks, for example, subjective complaints, around 50% of individuals typically meet the criterion for objective episodic memory impairment. The ‘cut-off’ for objectively defined cognitive impairment has typically been around 1.5 SD below education matched norms. Within this group with objective impairment, the rate of progression to dementia is around 12% per year [12], although this annual progression rate does not mean 100% of individuals will go on to develop dementia; a notable proportion will improve, remain stable, or not survive long enough to develop dementia [13]. A stricter ‘cut-off’ may result in higher progression rates of 16% per year [14].

A cardinal feature of AD dementia is amnesia and memory loss. Impairment to measures of episodic memory (i.e. the ability to learn and retain new information), has been identified as a reliable feature of AD and has been shown to correlate with the brain structures that are heavily involved in the AD process from its earliest stages [5]. Thus, episodic memory measures have become a standard method by which MCI populations have been categorized, and this has been described as amnestic MCI or aMCI, as distinct from impairment of other cognitive domains. This subtype of aMCI has been that most commonly associated with progression to AD dementia [12,15] and other subtypes of MCI are much less likely to progress to an AD type dementia [16].

Therefore, we see that a one-off assessment of cognition may be used to identify with some degree of sensitivity those at risk of developing a dementia, that is, at a ‘pre-clinical’ stage when used as part of a set of ‘diagnostic’ criteria. Furthermore, there may also be some degree of specificity for the type of dementia, with episodic memory impairment for example associated with AD dementia, whereas in Dementia with Lewy Bodies and Parkinson’s disease dementia, deficits to attention are one of the core cognitive symptoms [17,18]. Work in non-demented Parkinson’s disease patients has identified that MCI criteria were fulfilled in more patients for attention/executive deficits than amnestic deficits [19].

A number of cognitive tests have routinely been used to identify objective impairment of cognition in making the MCI classification. Several studies have been published which support both the sensitivity of particular measures and ‘cut-scores’ in identifying those who will progress to dementia and their specificity for specific types of dementia. Many of these studies have focused on AD dementia and the episodic memory domain, although there are numerous studies in dementia generally and which have assessed other cognitive domains.

The delayed recall part of the Logical Memory (LM) subtest of the Wechsler Memory Scale (LM-II) has been used in several clinical trials following aMCI populations longitudinally until they progress to a dementia. Two large studies using the same LM-II criteria will be cited here (the Alzheimer’s Disease Neuroimaging Initiative (ADNI) study [20] and the Alzheimer’s Disease Cooperative Study-Memory Impairment Study (ADCS-MIS [21])). In the ADCS-MIS, progression rate was 16%/year

---

**Figure 1**

<table>
<thead>
<tr>
<th>Disease Stage</th>
<th>FDA Approval</th>
<th>Subtle cognitive deficits alone</th>
<th>Increasing cognitive deficits</th>
<th>Detectable functional deficits</th>
<th>Dementia</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accelerated, based upon an effect on cognition</td>
<td>Standard, based on a single combined measure of cognition and function (e.g., CDR-SB)</td>
<td>Standard, based on coprimary measures of cognition and function or global rating</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Potential regulatory pathways in early Alzheimer’s disease: as the focus of drug development moves to earlier stages of Alzheimer’s disease, new guidance from the FDA suggests potential approaches to trial design that allow for regulatory flexibility and innovation. CDR-SB denotes Clinical Dementia Rating Sum of Boxes score.
for the first year and of the 214 participants progressing to dementia, 212 were classified with possible or probable AD. In ADNI, progression rate was 16.5%/year for the first year. Thus specificity for AD and progression rates from these trials, indicate LM-II may be particularly useful for predicting the development of AD dementia.

List learning tasks are seen by some researchers as having potential in the diagnosis of AD dementia at an early stage [22]. In addition, they may be easier to adapt for use in other languages and cultures than measures of paragraph learning (e.g. LM-II). Rabin et al. [22] identified a combination of California Verbal Learning Test (CVLT) and (WMS-III) LM measures as having the greatest predictive value for progression to AD over a 4 year period in an MCI cohort. An evaluation of the Rey Auditory Verbal Learning Test (RAVLT) for its value in predicting progression to dementia was conducted, in which 24 subjects (53.3%) developed dementia after being categorized with MCI, over a period of approximately three years [23]. The RAVLT measures of immediate and delayed recall (raw score, unadjusted for age and education) showed good sensitivity (>80%) though were similar to other measures (MMSE and IQCODE).

It has also been proposed that measures of memory that evaluate cuing may be more specific in the assessment of early AD dementia, where a profile of episodic memory impairment characterized by impaired free recall, unimproved or only marginally improved by cuing has been identified [24]. The Free and Cued Selective Reminding Test (FCSRT) is a multi-trial learning test using controlled learning and assessing Free Recall, Cued Recall, and Selective Reminding. In a study of 251 individuals diagnosed with MCI the FCSRT-IR scores of total recall, index of cuing free recall, free recall, delayed free recall, delayed total recall, and number of intrusions, all had area-under-the-curve values higher than 0.87 for predicting AD [26]. Importantly, age was a predictor of progression, but not education or sex; and the analyses identified cut-points for two age groups to maximize sensitivity and specificity of FCSRT measures.

Despite the sensitivity and specificity of these assessments, there are a number of caveats important to consider in their practical application to clinical assessment and the diagnosis of preclinical dementia. The use of a cut-off of 1.5 SD on measures of episodic memory to define aMCI has been shown to give an increased annualized progression rate to dementia versus the general population. However, other studies have used more or less strict cut-offs. It is likely that the use of a stricter cut-off will increase the likelihood of progression to dementia over a shorter period. However, this may then be at the expense of sensitivity in failing to identify those at an earlier stage of preclinical dementia. In addition, such cut-offs are on the basis of normative data for a particular measure and should ideally be age and education adjusted. Without such care, those with particularly high or low levels of ability may be misclassified. Therefore, the availability of normative data and also their relevance to the patient (e.g. culture, language, etc.) is of concern. Other tests have proposed cut-scores, different from a normative data based cut-off, as these have been determined by analyses to identify the score at which accuracy metrics are optimized. However, for many tests the application of such cut-off scores to wider populations has not been tested. Thus the cut-offs may be to some extent specific to the study population and in many cases have been studied in MCI type populations, that is, those already presenting with subjective complaints regarding cognition and/or some objective impairment. So again, applicability of cut-off scores more generally may be a concern. The data for episodic memory tests show that a number of measures give similar levels of accuracy in the identification of subjects likely to progress to dementia, and few studies have conducted head-to-head comparisons of tests when used in this way. Thus the use of published accuracy metrics is only part of the picture in determining which test to select, and the interpretation of such data and comparison between them presents many difficulties. A further concern with a single assessment of status is misclassification because of additional factors including lack of familiarity with test procedures, anxiety, fatigue, motivation and other transient influences on performance. Additionally, given the heterogeneity in the studies conducted, the application of any specific cut-off or cut-score on a measure provides very little information in determining the likely time course of the development of a dementia.

Given the challenges associated with the application of a single assessment of cognitive performance and the progressive neurodegenerative nature of the dementias, the use of these tests longitudinally to track decline may seem to be both a logical and a simple step. The use of a follow-up assessment to confirm cognitive decline requires neither a normative data cut-off to be applied. Furthermore, misclassification because of uncertainty regarding pre-existing ability and transient effects on performance should be avoided. However, in the next section we will outline the challenges that may be associated with the longitudinal assessment approach and the characteristics of suitable tests, some of which argue against the suitability of many of those tests most commonly used in single assessments of cognitive/neuropsychological status.

**Longitudinal assessment**

It is widely acknowledged that the cognitive decline which precedes AD and most other dementias occurs gradually. Thus the challenge clinically is how best to identify this deterioration. One solution is to administer tests longitudinally, the value of this approach being
Neuropsychological test data from the cognitively unimpaired population of the ADNI study over six years. Twenty-seven of the 256 subjects converted to aMCI an average of 2.5 years from study onset, and their data are presented separately from the 229 non-converters (controls). The scores are the mean (±SEM) over the study for Trail Making Part B and the Boston Naming test on the top row, and delayed recall scores for the Rey Auditory Verbal Learning Test and Logical Memory test on the bottom row.

Figure 2

Illustrated by several studies that have identified differing trajectories of performance in patients who subsequently went on to develop MCI to those who did not (e.g. [26]). Figure 2 presents data taken from the cognitively unimpaired population (n = 249) of the ADNI study, this group having a mean age of 76 years and an average MMSE of 29 at the outset. Over the course of 6 years, 27 individuals were diagnosed with aMCI, this occurring on average 2.5 years after the start of the study. The plots in Figure 2 show the data from four neuropsychological tests which were administered yearly during the study. It can be seen that for 3 of the measures, on study entry those who developed aMCI were already poorer than the non-converters at baseline. Whether this represents an earlier onset of decline, or was because of these patients always being inferior cannot be determined. For the delayed score of the LM test, a pattern of decline can be seen over the six years, though for the other three, this only became evident from year 3 or year 4 onwards. This pattern though is influenced by the training effects seen in non-converters on each measure, and if difference scores between the two groups are calculated, the declines can be seen to occur before the average time of conversion to aMCI. Although cognitive decline may be rapid in some forms of dementia, for example, early onset AD [20], in the majority of cases the decline is much more gradual. In addition, rate of decline has often been shown to be disease severity dependent, with smaller rates of cognitive decline at the earliest stages of dementia [27]. However, this pattern may also be in part measure dependent, in that different measures of cognition may show differential sensitivity at different stages of the disease [28,29].
Thus the detection of longitudinal decline in preclinical dementia may require the identification of very subtle changes, and distinguishing such change may present a number of challenges, including those of appropriate test selection. As in any area of cognitive assessment, basic psychometric properties, particularly floor and ceiling effects are critical, but a number of other considerations unique to longitudinal assessment also come into play.

An important factor to consider in the detection of the changes in cognitive function which precede dementia is that normal ageing is associated with a steady decline in most domains of cognition from the twenties onwards. This pattern has been investigated in an extensive series of studies by a research group led by led by Timothy Salthouse at the University of Virginia, USA [30**]. The group has assessed thousands of healthy individuals across the adult age range using a wide variety of neuropsychological tests. The consistent finding from the research has been for linear declines to be present from the twenties onwards on a range of measures of attention, information processing, reasoning, spatial skills and various aspects of memory; only measures of the use of vocabulary showing a different pattern.

Similar patterns of decline with normal ageing have been identified using a set of computerized tests of major domains of cognitive function [31,32]. An example of these steady declines for a measure of focused attention and information processing is presented in Figure 3. These data were gathered over 18 months between 2010 and 2011 from over 90 000 individuals who logged onto a website and performed 12 min of cognitive testing. As with the data from the Salthouse programmes, the declines are linear across the age range and begin in the late twenties. Further, the error bars are confidence intervals and thus up to the 70s the declines are significant for each successive five year age band.

In the new preclinical AD guidelines, Sperling et al. recognize that at the earliest stages the changes to cognitive function in patients will be subtle compared to their own baseline, but will exceed that expected from normal ageing, while not meeting the criteria for MCI [8**]. It is unlikely that a cross-sectional test could help identify such an increased rate of normal ageing. The normal distribution ensures a wide range of abilities, and considerable movement within a normal distribution could still leave the individual in the normal range. For preclinical AD, longitudinal assessment thus appears to be the only feasible methodology with appropriately sensitive tests to identify a change in function which can be considered pathological. For MCI, the new guidelines make it clear that if repeated assessments of cognitive function are available, then a decline in performance should be evident over time [7**,25]. Although such repeated assessments may not be essential for the initial diagnosis in MCI, subsequent decline is essential to confirm the original diagnosis.

While longitudinal assessment of cognitive function will be essential to capture the increased rates of decline which occur in preclinical dementia, and will have great utility in prodromal AD; repeated testing of cognitive function has a number of methodological difficulties. These include what are termed ‘training’ or ‘practice effects’, where the quality of test performance may change over repeated sessions because of factors unrelated to the cognitive status of the individual. For example Salloway et al. found learning effects on the Alzheimer’s Disease Assessment Scale, cognitive subsection (ADAS-cog [33]) in a clinical trial in MCI [34]. Such training effects can also occur in AD patients, and may be responsible for the failure to detect expected rates of cognitive decline in placebo group in some trials (e.g. [4*,35]). As the successful AD compounds have been found to slow the rate of cognitive decline [1**,2], this failure to see decline under placebo generally results in a negative outcome for the novel treatment; leaving the field unsure whether the compound was potentially active or not. Short intervals between testing can be a problem, as can frequent testing, as many neuropsychological tests do not have alternate forms [36**], the performance of the test improving with re-testing due to the participant being able to remember for example the stimuli used in memory testing. Good practice in this area includes pre-baseline familiarization with the tests, as well as the employment of tests which have alternate forms and which have been established to be relatively free of [37] ‘practice effects’.

**Figure 3**

Illustration of cross-sectional declines in normal ageing for a measure of focused attention and information processing gathered recently from individuals who logged onto a website and performed 12 min of cognitive testing.

(Taken from [32]).
With appropriate design and standardization, tests can be developed which properly characterize the quality of performance over repeated assessments [37]. In one study an automated test of paired associate learning identified clear deterioration over a two year period in subjects with ‘questionable dementia’ (QD) as well as Alzheimer’s patients, but not in normal controls [38]. All those with QD who showed decline on the measure went on to be diagnosed with AD. In another study, AD diagnosis was made using an algorithm on the basis of decline over a three year period on a memory test (Neuropsychological Assessment Battery List Learning). Categorization to the AD diagnosis using this algorithm showed a 73% increased risk of a consensus diagnosis of AD at follow-up [39]. This is one of several studies which demonstrate that decline on cognitive tests can be applied as an accurate diagnosis of likely progression to a dementia.

Automated versus non-automated tests
Tests widely employed in human Cognitive Neuroscience have only made a limited incursion into clinical trials in dementia, or any CNS field for that matter; the two fields having developed along largely independent paths. In general the tests used in CNS clinical trials have either been developed by medically qualified physicians such as the Mini-Mental State Examination [40] or are pencil and paper tests typically used by clinical neuropsychologists. The latter case is exemplified the ADASCog which has been (and still is) the most widely employed outcome measure in therapeutic dementia trials [33]. One problem is that few neuropsychological tests were developed specifically for repeated use, and thus as mentioned earlier, have limited parallel forms; which can lead to training effects which mask possible treatment effects. The populations for preclinical AD trials will be normal individuals, and traditional neuropsychological tests have generally performed poorly in this population with repeated testing. Training effects because of repeated testing have been identified in a range of tasks, particularly neuropsychological tests [41–43], even when the gaps between testing are as much 2.5 years [43]. The ADNI study includes a non-cognitively impaired cohort of over 250 individuals aged 60–90 who are being tested annually on a wide range of traditional non-automated neuropsychological tests. Analysis of these data identified just such training effects on the various neuropsychological tests employed, which have lasted up to six years on some measures [44]: having effect sizes that generally exceeded those identified with registered antidementia drugs [2]. Furthermore, none of the tests was able to detect any consistent pattern of decline over the six year period, despite 11% of the sample having converting to MCI. The industry sponsored therapeutic trial of over 5000 preclinical AD patients mentioned earlier has solely employed such neuropsychological tests as the primary measure of change over the study [10]. However, the A4 study also mentioned earlier has not followed this path, and has included computerized cognitive tests in the assessment profile.

The general reluctance of the pharmaceutical industry to employ computerized tests of cognitive function in CNS trials is difficult to understand. It is hardly the case that such methods are new, untested or even innovative. For example in 1868 the researcher Francis Donders developed automated tests of simple reaction time (SRT) and choice reaction time (CRT). He demonstrated that reaction times were longer in CRT [45], reflecting the extra information processing necessary to perform CRT compared with SRT, a finding which remains one of the central tenets of cognitive psychology. Since then, these tests together with other assessments including vigilance tasks have been the mainstay of attention testing in cognitive neuroscience [46**].

Possibly the persistence shown by the pharmaceutical industry in retaining traditional neuropsychological testing despite the manifest limitations which have been identified can be excused by the absence of clear and authoritative guidance for clinical trials in dementia. However, a Position Paper published in 1997 by the International Working Group on Harmonisation of Dementia Drug Guidelines was extremely clear on this matter [47]. The Working Group proposed ‘Uniform Criteria/Requirements for Optimal Cognitive Assessment Instruments’:

‘There was agreement regarding the test characteristics necessary for a measure to be suitable for an AD trial. These included validity (the instrument must measure the intended disease-relevant cognitive functions); reliability (test–retest, also inter-rater and intra-rater if scoring is subjective); appropriate sensitivity range (absence of ceiling and floor effects, taking into account anticipated decline over the duration of the trial); availability of longitudinal data (information should be available on expected change over the course of the trial, which is helpful for making power estimates); information on practice effects (determines the need for practice sessions before starting a trial); and availability of equivalent forms (for repeated testing over the course of the trial).’ (Page 35).

On computerized testing, the Group wrote the following:

Computerized Testing
‘Computerised procedures currently are used extensively in general psychopharmacology, and some systems have been developed specifically for use with demented patients. There is evidence that, after an initial familiarization, properly implemented computerized procedures can be perfectly acceptable to AD patients (Ferris et al., 1988; Simpson et al., 1991). 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 gray areas involving differences in interpretation. These advantages can reduce variability both from session to session for a patient, and also between different national and international sites. Automated procedures recently have been shown to be more sensitive than the standard tests that are used extensively in this field (Mohr et al., 1996), and the sensitivity to anticholinesterases in patients with AD also has been established (Siegfried, 1993). 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 (e.g. Wesnes et al., 1987; Nicholl et al., 1995). Some tests of attention such as vigilance can be run only on computers. However, before being used in major trials, extensive assessments of reliability, validity, and utility of these tests must be made. In addition to the aforementioned basic test criteria, crucial requirements for automated testing include the recording of responses via simple response buttons or touch screens, not the keyboard; absence of unwanted negative feedback from the tasks; timing routines that are accurate to the nearest millisecond and that are made independently of the internal clocks of the computer; established reliability of software and hardware; presentation of information using specially constructed fonts that are clearly visible to patients with AD, and security of automatically recorded data files that can be accessed only by authorized site staff members. The Work Group concluded that computerized procedures initially should be used together with the established procedures in the field (e.g. the ADAS) so that the comparable utility and sensitivity of the two types of testing can be identified. If clear advantages of computerized procedures are demonstrated, such procedures might supersede existing methods.’ (Page 36).

It should be noted also that one of the authors of the paper helped to develop the ADAS-cog. Nonetheless the pharmaceutical industry has largely ignored these guidelines, and the vast majority of the AD trials conducted to date have used the ADAS-cog as the primary outcome. The industry has even compounded its problems by employing the more recently developed Neuropsychological test Battery (NTB) for Alzheimer’s trials [28]; although it is composed entirely of neuropsychological tests, including some of those employed in ADNI. The risk of training effects either resulting in improvements or a lack of decline on the NTB in placebo treated patients cannot therefore be discounted. Unfortunately just such effects were seen in one of the few successful AD trials in recent years [48°]. This double-blind, placebo controlled clinical trial in mild AD, saw training effects on the NTB in the placebo group which at three months obscured the possible benefits of a medical food which had previously been established to be effective in AD treatment. Fortunately, despite the training effects still being present at 6 months in the placebo group, the active treatment just edged ahead of placebo at this time [48°]. This was despite a major mechanistic target of the medicinal food (the plasma levels of the cognitively toxic substance homocysteine) showing a significant decline over the period on active treatment compared with a significant increase under placebo. The authors wrote: ‘This may be a genuine placebo effect, or partly because of other factors such as learning/familiarity with the tests themselves, which again given the mildly affected state of the patients may have happened. In future studies, using two baseline scores and parallel versions of tests may overcome this’. (Page 92).

Translational medicine is crucial in drug development, and ideally the same tests of cognitive function could be used throughout the development process from Phase I to Phase IV. The 1997 Working Group [47] report contained a section entitled ‘Relevance of Stage of Drug Development for the Cognitive Measures Selected’: ‘There can be differences in the types of tests or numbers of tests that are used, depending on the stage of development of the drug. A single comprehensive measure such as the ADAS is adequate and appropriate for phase III pivotal trials, but the Work Group concluded that more detailed cognitive testing sometimes might be appropriate and helpful in early phase II trials when the nature of a drug’s activity is uncertain. A broad cognitive testing strategy early in clinical development might enable the profiling of a novel drug in early phase trials. This early profiling could lead to selection of an optimally sensitive battery for use in phase III trials.’ (Page 36)

Phase II is a crucial time for an antidementia drug as the optimal dose range must identified [1°]. Unfortunately the ADAS-cog requires very large samples which has led some companies to move directly from Phase I to Phase III; with has been a possible reason behind some recent failures of antidementia drugs. A more appropriate practice in Phase II as suggested above is to use a more sensitive test battery in order that the dose range can be properly explored in smaller groups of patients. Another of the very few recent successes of a therapy with a novel mechanism of action was reported in 2013. This was with ORM-12741, a highly potent and selective alpha-2C adrenoceptor antagonist, which was studied in a 12 week Phase IIA trial in moderate AD patients. The study used a computerized battery designed for repeated administration as the primary endpoint [49] and identified potential efficacy in the three armed trial in a total of 100 patients. Computerized tests of working and episodic memory showed a statistically reliable benefit over the 12 weeks with both doses of the compound, preventing
the decline seen over the study under placebo. The usual practice in this area involves treatment periods of six months or longer, and one consequence of this shorter dosing period was that it enabled the trial to be completed more quickly, an important advantage in Phase II. Further, significant reductions in caregiver distress were seen on the Neuropsychiatric Inventory, adding independent support to the cognitive benefits.

Translational medicine can also be employed in Phase I, using a model involving the administration of scopolamine to volunteers. This procedure produces a temporary profile of cognitive deficits which closely mirrors that seen in AD, and has been widely employed for over 25 years [50,51**]. The scopolamine model can not only help identify compounds likely to work in AD patients, but it also enables the dose-response range to be assessed. Curvilinear dose-responses are common with cognition enhancers [1**] and if these are identified in the model, the developer will know that care must be taken to determine the optimal dose in Phase II. Alternatively if the dose-response appears linear, in Phase II the developer can with more confidence administer the compound at levels approaching the highest tolerated dose.

Adaptive designs are becoming more widely used in dementia research in recent years (e.g. [52]). In such studies the response to treatment is monitored regularly, and doses can titrated in patients, or dropped for ineffectiveness, and others selected until an optimal dose is identified. Here sensitive cognitive tests are extremely valuable as they enable the early identification of any beneficial effects in relatively small groups. The FDA has also in 2013 issued for comments a draft Guideline for Industry entitled ‘Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products’ (http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM332181.pdf). This statement captures the relevance of such procedures to this research area:

‘Prognostic enrichment strategies are also applicable, or potentially applicable, to the study of drugs intended to delay progression of a variety of diseases, such as AD, Parkinson’s disease, rheumatoid arthritis, multiple sclerosis, and other conditions, where patients with more rapid progression could be selected; . . .’ (page 5)

A final question is whether computerized cognitive tests could prove more useful than their non-automated counterparts in trials in prodromal and preclinical AD? A trial has recently shown computerized assessments to detect improved cognitive function in MCI patients administered nicotine patches for six months [53**]. Further, in a five-year clinical trial, 257 hypertensive, but not cognitively impaired older adults (mean age 76 years), were treated with an angiotensin II-receptor blocker or placebo [54], and a computerized cognitive battery plus traditional neuropsychological tests were administered yearly. It was found that active treatment statistically significantly reduced the rate of decline seen under placebo over the study on computerized measures of episodic memory and attention, but not on the pencil and paper tests. The effect sizes of the benefits were comparable to those seen with anticholinesterase therapy in AD [2]. The overall methodology of cognitive testing in this study was comparable to that which is (and will be) used in preclinical AD; suggesting that computerized testing could prove more appropriate in such trials than neuropsychological tests.

The status of cognitive testing in other CNS disorders

Other fields of CNS drug development besides dementia have similar problems with insensitive and/or impractical cognitive outcome scales; two examples being the Multiple Sclerosis Functional Composite (MSFC) and the Glasgow Coma Scale (used in trials in stroke and traumatic brain injury). Cognitive dysfunction in schizophrenia has been targeted as a major target for treatment. This led to the NIMH Division of Mental Disorders, Behavioral Research, and AIDS sponsored the Measurement and Treatment Research to Improve Cognition in Schizophrenia programme (MATRICS; www.matrics.ucla.edu). This was a broad-based consensus process that included the academic community, NIH, FDA, and the pharmaceutical industry; and a fantastic example of collaboration between various bodies which should become a model for many CNS areas. A group of internationally respected, sincere and highly motivated individuals worked tirelessly for years to establish consensus agreement on the full profile and extent of cognitive impairment in schizophrenia. They then determined by a rigorous and egalitarian consensus process what would be the ideal tests to serve as the outcome measures in the clinical trials for compounds to treat this clearly unmet medical need. The outcome was the MATRICS Consensus Clinical Battery (MCCB); which involves 10 tests/tasks; only one of which is automated, the others being widely used neuropsychological tests [55]. Sadly, in various trials of compounds to treat the cognitive impairment associated with schizophrenia, notable training effects were seen on the MCCB, which in some cases have masked potential treatment effects (e.g., [56,57]). Regarding the logistics of clinical trials there are various other potential limitations of the MCCB which include the full battery taking 60–90 min to complete (www4.parinc.com/Products), the need for highly specialized administration and test scoring, and that initially it was only available in English (though there are now over 10 language versions). Further, the selection criteria for the tests apparently did not include established drug sensitivity. These ‘limitations’ are now being addressed and revisions to the MCCB and then methodology behind its use are well
underway. As mentioned earlier, this massive project has redefined how instruments should be developed; but nonetheless by using widely used neuropsychological procedures, still did not achieve the initial impact in drug development it truly deserved. Many other CNS fields are also beginning to explore new clinical targets (see [58**]). To properly meet the needs of multiple groups of patients with cognitive dysfunction — it is crucial that the selection and/or methodological development processes of the outcomes are cognizant of the various hard lessons learned over the last 25 years (or more); because it is these instruments which will determine the eventual success of future therapy.

**Conclusion and future perspectives**

Cognitive testing already plays a major role in the identification of MCI using one-time administration assessments, and can help identify individuals likely to develop various types of dementia. However, the broad applicability of some tests and the possibility of misclassification are major limitations of this form of cross-sectional assessment. Longitudinal assessments should improve the reliability of diagnoses in MCI, and will almost certainly prove to be the only viable technique for trials in preclinical dementia. Besides this, in any therapeutic CNS trial, repeated cognitive assessments will always be essential; and instruments which can reliably detect change in function in a number of core cognitive domains will confer a significant advantage to such work. However, challenges can exist in the interpretation of repeated cognitive assessments when measures are employed which were not specifically designed for this purpose. Certainly high sensitive and reproducible tests with be crucial in the differentiation of pathological cognitive decline from normal ageing. Although automated testing is currently under employed in therapeutic clinical trials, it has still been used widely in dementia research and for many decades (e.g. [59–61]) and in the prodromal stages has been making significant inroads in this area, as is evidenced by a recent review of 16 test systems (mostly computerized) which could be suitable for identifying the early signs of MCI [36**]. Examples from the 1990s of positive effects of antiedementia drugs being detected by computerized testing in Phase I [62,63], replicated in small Phase II trials [64,65] and then confirmed in larger Phase III studies using traditional tests [66,67], exist for the tacrine analogue, velacrine, and the vasopressin agonist, S-12024. The recent example of the successful Phase II trial of the alpha-2C adrenoceptor antagonist plus other promising Phase II findings just reported with the histamine H3-receptor antagonist, SAR110894 [68], show the enduring value of using sensitive cognitive assessments in early Phase development. Computerized tests can assess a broader range of aspects of cognition which are particularly relevant to the deficits in AD and other dementias, as well as in prodomes for these diseases; and can also facilitate the ease of test administration and the quality of data capture. Furthermore, many cognitive tests can be administered over the internet, and this methodology could prove particularly useful in large long-term studies in preclinical and prodromal dementia.

**References and recommended reading**

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


  Could be viewed as the moment the future of research in dementia turned to prevention and trials in healthy individuals.


14. Petersen RC, Thomas RG, Grundman M et al.: Vitamin E and
15. Petersen RC, Roberts RO, Knopman DS et al.: Mild cognitive
value of CSF markers of Alzheimer’s disease pathology in
patients with subjective cognitive impairment or mild
cognitive impairment in the DESCRIPA study: a prospective
17. Emre M: Dementia associated with Parkinson’s disease.
19. Caviness JD, Driver-Dunkley E, Connor DJ et al.: Defining mild
cognitive impairment in Parkinson’s disease. Mov Disord 2007,
22:1272-1277.
20. Petersen RC, Aisen PS, Beckett LA et al.: Alzheimer’s Disease
Neuromaging Initiative (ADNI): clinical characterization.
Neurology 2010, 74:201-209.
21. Schneider LS, Raman R, Schmitt FA et al.: Characteristics and
performance of a modified version of the ADCS-CGIC CIBIC+
for mild cognitive impairment clinical trials. Alzheimer Dis
22. Rabin LA, Pare N, Saykin AJ et al.: Differential memory test
sensitivity for diagnosing amnestic mild cognitive impairment
and predicting conversion to Alzheimer’s disease.
Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2009,
16:357-376.
23. Isella V, Villa L, Russo A, Regazzoni R, Ferrarese C, Appollonio IM:
Discriminative and predictive power of an informant report
in mild cognitive impairment. J Neurol Neurosurg Psychiatry 2006,
77:166-171.
24. Ivanov A, Adam S, Van Der Linden M et al.: Memory evaluation
with a new cued recall test in patients with mild cognitive
25. Sarazin M, Barr C, De Rotru J et al.: Amnestic syndrome of the
medial temporal type identifies prodromal AD: a longitudinal
26. Howieson DB, Carlson NE, Moore MM et al.: Trajectory of mild
cognitive impairment onset. J Int Neuropsychol Soc 2008,
14:192-198.
27. Van Der Vlies AE, Koedam EL, Pijnenburg YA, Twisk JW,
Scheltens P, Van Der Flier WM: Most rapid cognitive decline in
APOE epsilon4 negative Alzheimer’s disease with early onset.
28. Harrison J, Minassian SL, Jenkins L, Black RS, Koller M,
Grundman M: A neuropsychological battery for use in Alzheimer
29. Cano SJ, Posner HB, Moline ML et al.: The ADAS-cog in
Alzheimer’s disease clinical trials: psychometric evaluation of
the sum and its parts. J Neurol Neurosurg Psychiatry 2010,
81:1363-1368.
University Press; 2010.
31. Cano SJ, Posner HB, Moline ML et al.: The ADAS-cog in
Alzheimer’s disease clinical trials: psychometric evaluation of
the sum and its parts. J Neurol Neurosurg Psychiatry 2010,
81:1363-1368.
University Press; 2010.
33. Rosen WG, Mohs RC, Davis KL: A new rating scale for
34. Salloway S, Correia S, Richardson S: Key lessons learned from
short-term treatment trials of cholinesterase inhibitors for
35. Gold M: Study design factors and patient demographics and
their effect on the decline of placebo-treated subjects in
randomized clinical trials in Alzheimer’s disease. J Clin
36. Snyder PJ, Jackson CE, Petersen RC et al.: Assessment of
cognitive impairment in mild cognitive impairment: a comparative study.
37. Wesnes K, Pincock C: Practice effects on cognitive tasks: a
associate performance in the early detection of DAT. J Int
decline and conversion to Alzheimer’s disease in older adults
using the NAB list learning test. J Int Neuropsychol Soc 2010,
16:651-660.
40. Folstein MF, Folstein SE, McHugh PR: ‘Mini-mental state’. A
practical method for grading the cognitive state of patients for
41. McClelland GR: The effects of practice on measures of
42. Wilson RS, Beckett LA, Barnes LL et al.: Individual differences in
rates of change in cognitive abilities of older persons. Psychol
43. Salters TA: Influence of age on practice effects in
longitudinal neuropsychological change. Neuropsychology 2010,
24:563-572.
44. Wesnes KA, Schneider LS: Are neuropsychological tests such
as those used in ADNI suitable for long-term trials of cognition
enhancers for preclinical Alzheimer’s disease? J Nutr Health
Aging 2012, 16:810.
45. Donders FC: Die schnellheit psychischer prozesse. Arch für
Anatomie und Physiologie und wissenschaftliche Medizin
46. Wesnes KA, Brooker H: Pharmacodynamic evaluation: CNS. In
Edited by Vogel HG, Maas J, Gebauer A. Berlin Heidelberg:
47. A detailed and extensive review of the full range of outcomes
currently used in drug research.
48. 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 Harmonisation of Dementia
49. Scheltens P, Twisk JW, Blesa R, Scarpini E, von Arnim CA,
Bongers A, Harrison J, Swinkels SH, Stam CJ, de Waal H,
Wurtman RJ, Wieggers RL, Vellas B, Kamphuis PJ: 1’ Efficacy of
Sovenan in mild Alzheimer’s disease: results from a
A stunning example of an almost failed trial of a really impressive medical
food which was badly let down by the utilization of inappropriate neuropsychological
outcome methodology.
50. Wesnes KA, Revell A: The separate and combined effects of
scopolamine and nicotine on human information processing.
Psychopharmacology 1984, 84:5-11.

A good review on the current status of the scopolamine model with novel data from preclinical and clinical trials demonstrating the utility of the model for dementia.


A successful trial in MCI — one of the very few.


A fabulous and detailed summary of the new approach to drug development in CNS, a real paradigm shift.


