International Psychogeriatrics (2007), 19:3, 383–389 © 2007 International Psychogeriatric Association
doi:10.1017/S1041610207005169 Printed in the United Kingdom

CONSENSUS CONFERENCE PAPER

Improving dementia assessment by reducing sample heterogeneity

Marshal Folstein

Professor of Psychiatry at Tufts University School of Medicine, Boston, Massachusetts and Johns Hopkins Medical Institutions, Baltimore, Maryland, U. S. A.

ABSTRACT

Outcome assessment for clinical dementia trials could be enhanced by using the Mini-mental State Examination (MMSE) and the Hopkins Verbal Learning Test (HVLT), in addition to imaging and genetic screening. The statistical power of trials could be increased if heterogeneity of the sample were reduced by the administration of a risk factor inventory which could be used for subject selection or selection of drug responders.

Key words: outcome assessment, clinical dementia trials, Mini-mental State Examination, Hopkins Verbal Learning Test

Two problems prompt reconsideration of outcome measures used in dementia clinical trials. First, drugs for dementia help a few patients greatly, but most experience variable, small changes in cognition over long periods of time. Inappropriate measurement is an unlikely explanation for these results. The apparent small and variable response could be due to either ineffective drugs or to the heterogeneity of the samples. Second, statistically significant results based on current outcome measures do not convince public officials that the drugs are helpful enough to warrant their cost. The officials may also be put off because technically excellent outcome measures like ADAS-Cog are not readily recognized by the public and clinicians. Translation of research results into practice requires persuasion which is enhanced by the use of transparent methods that, on the face of it, are clinically relevant. The purpose of this conference hopefully is to begin to solve these problems.

Assessment of improvement of dementia must include tests of memory and other cognitive functions which define the syndrome of dementia. Core symptoms of Alzheimer's disease (AD) which emerge during its course include

Correspondence should be addressed to: Professor M. Folstein, 11 Hamlet Hill Road, Unit 406, Baltimore, MD 21210, U.S.A. Phone and Fax: +1 410 433 6109. Email: mfolstein@rcn.com. Received 30 Oct 2006; accepted 31 Oct 2006. First published online 26 March 2007.

https://doi.org/10.1017/S1041610207005169 Published online by Cambridge University Press

memory, language, and visual spatial and executive function (Folstein and Whitehouse, 1983). Even though I have a conflict of interest, I recommend that improvement of assessment of the dementia of AD be demonstrated by a Mini-mental State Examination (MMSE) increase of three points or greater, over one year. The MMSE is well known by clinicians and by research workers who know that the expected decline of MMSE scores for AD patients is two to four points per year (Folstein *et al.*, 1975; Anthony *et al.*, 1982; Crum *et al.*, 1993; Aguero-Torres *et al.*, 1998; Burback *et al.*, 1999; Han *et al.*, 2000). Since the standard deviation of MMSE in most studies is 3, an increase of three points per year represents 2 standard deviations from the expected outcome without treatment and is thus not likely to occur by chance.

Another outcome that I would readily accept as demonstrating improvement of the symptoms of AD would be a statistically significant increase in a test of delayed recall, because memory loss is a defining feature that is also well known to the public. There are many suitable memory tests, but one I would recommend would be the Hopkins Verbal Learning Test (Rasmusson *et al.*, 1995). It is reasonably brief, tests several aspects of memory, and is reliable. The MMSE combined with the HVLT gives reasonable sensitivity and specificity and is brief and inexpensive. The capacity to detect mild dementia indicated by a MMSE cutting score between 25 and 26 was 88% sensitive and 93% specific. The HVLT, with a cutting score between 18 and 19 for sum of recall trials, was 96% sensitive and 80% specific (Frank and Byrne, 2000).

Currently most clinical trials of AD drugs use the MMSE for inclusion criteria of severity of dementia and the ADAS-Cog for measuring change of cognition. The MMSE and ADAS-Cog assess similar cognitive domains. The Alzheimer Disease Assessment Scale is a battery of tests of memory, orientation, language and praxis (Mohs and Cohen, 1988). Memory and language are emphasized. Another frequently used measure is the Clinical Dementia Rating (CDR), a subject and informant-based assessment which rates severity, but does not directly test cognitive domains and activities of daily living (Morris, 1993).

Another indicator of the severity of the dementia syndrome is the performance section of intelligence tests. Most of these intelligence subtests assess what is now called executive function. There has been increased interest in tests of executive function, especially since the work of Royall who devised a short battery of tests of executive function (Royall *et al.*, 1992). In the MMSE, items for registration and serial 7 s assess executive function. However, consensus has not yet been reached as to the necessary and essential features of executive function. Albert has shown that a combination of a memory test and a test of executive function after a three-year follow-up would predict cognitive decline in normal subjects and subjects with mild cognitive impairment (Albert *et al.*, 2001). However, even without adding new measures, efficiency and cost of clinical trials would

https://doi.org/10.1017/S1041610207005169 Published online by Cambridge University Press

be improved if outcome measures were shorter and required less training. Computer-based testing of subjects, which would increase reliability and reduce cost of administration, is in the early stages of development, but will require wide use before all its limitations are known. Computer-based, telephone-based (Brandt *et al.*, 1988) and video-based measures (Karweit *et al.*, 1994) are in development. Other outcomes to be measured include noncognitive symptoms (Reisberg *et al.*, 1996; Cummings, 1997) and reduction of social competency (Janofsky *et al.*, 1992; Bassett *et al.*, 1998).

The power of currently used outcome measures would be increased by reducing the heterogeneity at the levels of symptoms, pathology, and etiology of the samples accepted for clinical trials. I first became aware of heterogeneity of response of dementia symptoms to drugs while testing physostigmine as an esterase inhibitor. In the series we published in 1991, one individual out of 13 improved dramatically within minutes of intravenous injection. Before injection he could not copy a cube, but afterwards he could. The same individual showed a marked increase in cerebral perfusion measured by PET scans. Increased perfusion was associated with cognitive improvement. Other cases improved but not nearly as dramatically as the first case (Tune *et al.,* 1991). The dramatic variability in response was not due to the reliability, validity, sensitivity or specificity of symptom measures, or to the capacity to demonstrate pathology by functional imaging techniques.

There are several known sources of response variability and probably many unknown ones. One of the most important is the one I call "the fallacy of the early case." There is an understandable desire to treat AD early on in the course and recently there has been a trend to treat precursors of dementia such as "minimal cognitive impairment." Unfortunately, these good intentions founder on the rocks of our diagnostic capability. It has been known for 30 years that 30% of cases diagnosed early in the course of illness do not show the characteristic AD pathology (Nott and Fleminger, 1975). More recently, mild cognitive impairment (MCI), which is sometimes the precursor to cognitive deterioration, has been studied pathologically. In the words of the authors, "The neuropathological outcome of amnestic MCI following progression to dementia is heterogeneous . . . it includes AD at a high frequency . . . but complex neuropathologic findings including 2 or more distinct pathologic entities contributing to dementia may be common in community-based cohorts. Neither demographic variables nor cognitive measures had predictive value in determining which patients diagnosed with MCI will develop the neuropathologic features of AD" (Jicha *et al.*, 2006). It may be that the price of early treatment is currently too high to justify the practice.

Sources of heterogeneity of samples of AD cases include variations of synaptic loss as well as the number and distribution of plaques and tangles (Terry *et al.*,

https://doi.org/10.1017/S1041610207005169 Published online by Cambridge University Press

1991; Samuel *et al.*, 1994). Many pathways in the pathogenesis of AD have been identified which include amyloid build-up and clearance, synaptic plasticity inflammatory response and transmitter pathway vulnerability. Since different drugs will affect different pathways, a mixture of pathogenic types will decrease the power for any given drug trial to detect improvement.

Heterogeneity is introduced by inaccurate diagnosis. Many subjects of clinical trials do not have AD. Among cases diagnosed as AD, 10–15% have fronto-temporal dementia, Lewy body disease, Parkinson's disease, hippocampal sclerosis, or cerebrovascular disease (Alexander *et al.*, 1995; Cercy and Bylsma, 1997; Hatanpaa *et al.*, 2004).

In addition to neuropathological heterogeneity, variation is seen in disease risk factors, course, and symptoms. The variation in risk factors is both genetic and environmental. Genetic variation – APOEϵ4 and mutations of chromosomes 1, 14, 21 and probably 12 and 9 – may well influence response to treatment (Evans *et al.*,1997; Pericak-Vance *et al.*, 1997; Hardy and Gwinn-Hardy, 1998; Gatz *et al.*, 2006). Environmental risks which should be considered include head trauma (Lye and Shores, 2000), years of schooling, probably premorbid intelligence (Valenzuela and Sachdev, 2006) and possibly aluminum (Flaten, 2001). Diabetes (Arvanitakis *et al.*, 2004; Devanand *et al.*, 1996; Qiu and Folstein, 2006), and cardiovascular disease (Hayden *et al.*, 2006) appear to increase the incidence of AD in portions of the population.

Variation in course of illness is related to risk factors such as age, gender, education, comorbidity, and functional disability which predict survival (Aguero-Torres *et al.*, 1998). It would seem reasonable to designate cases as rapid decliners versus slow decliners. This could be assessed by dating onset of illness from the first signs of memory loss noted by the family. Since the average course of decline of the MMSE is known to be two to four points per year, cases can be designated according to whether they decline more or less than the average. The cause of variation of course of illness includes comorbid vascular disease or some other complicating pathological state such as Lewy body disease. In my experience some families with the disease have a more benign course than others.

Clinical symptoms, which occur in some cases but not others, also affect outcome. The presence of noncognitive symptoms predicted a more rapid course in a study by Stern *et al.* (1987). This effect is possibly due to the inclusion of cases with Lewy body disease in the sample since this disorder has a high frequency of noncognitive symptoms.

It is likely that these multiple pathways to AD will be related to differential responses to treatment because different etiological mechanisms produce different neuropathology.

Because of the heterogeneity in cases that are currently accepted in clinical trials, improvement of measurement accuracy over current levels will bring only

https://doi.org/10.1017/S1041610207005169 Published online by Cambridge University Press

modest improvement of overall accuracy. For example, most evaluations of instrument accuracy are made using clinical diagnosis as the standard. If clinical diagnosis predicts pathology with 85% accuracy, a measure that is 85% accurate in detecting clinical diagnosis will only be 73% accurate in detecting pathology. To continue this game, improving instrument accuracy against clinical diagnosis from 85 to 90% would yield only a 3% improvement in overall accuracy. (In some studies instrument validity has been demonstrated against neuropathology However, these studies seldom demonstrate the accuracy of the instrument used to measure the level of severity of cases entering trials which would have occurred many years before autopsy.) It is my opinion that improvement in accuracy of outcome measures will come from the combination of current measures with measures not yet in common use in clinical trials: scans, plasma and CSF levels of tau and Aβ42 and genotypes that index the various types of pathology and etiology. A new questionnaire is needed which records the presence of all known risk factors for each case, so that either selection for entry or analysis of outcome can account for heterogeneity as a source of response. Items to be included in a risk factor heterogeneity questionnaire: age and age of onset, education/IQ/cognitive reserve, gender, family history, APOEϵ4 and other genes, head trauma, depression, diabetes and cardiovascular risks, rates of decline and noncognitive symptoms. Sources include patient, family and medical records. The score could be determined by counting the number of risk factors as well as categorizing cases according to specific risk factors such as age of onset.

Since clinicians use approved drugs for heterogeneous groups of patients, large effects and large sample sizes will be needed in trials so that the drug is seen by clinicians to be effective in their practice. Currently used measures like the MMSE are adequate to demonstrate these large differences.

Conflict of interest declaration

Marshal Folstein receives royalties from the publisher of MMSE and TICS. His salary is supported by NIA, AF-21790.

References

Aguero-Torres, H., Fratiglioni, L., Guo, Z., Viitanen, M. and Winblad, B. (1998). Prognostic factors in very old demented adults: a seven-year follow-up from a population-based survey in Stockholm. *Journal of the American Geriatrics Society*, 46, 444–452.

Albert, M. S., Moss, M. B., Tanzi, R. and Jones, K. (2001). Preclinical prediction of AD using neuropsychological tests. *Journal of the International Neuropsychology Society*, 7, 631–639.

Alexander, G. E., Prohovnik, I., Sackeim, H. A., Stern, Y. and Mayeux, R. (1995). Cortical perfusion and gray matter weight in frontal lobe dementia. *Journal of Neuropsychiatry and Clinical Neuroscience*, 7, 188–196.

https://doi.org/10.1017/S1041610207005169 Published online by Cambridge University Press

Anthony, J. C., LeResche, L., Niaz, U., von Korff, M. R. and Folstein, M. F. (1982). Limits of the "Mini-Mental State"as a screening test for dementia and delirium among hospital patients. *Psychological Medicine*, 12, 397–408.

Arvanitakis, Z., Wilson, R. S., Bienias, J. L., Evans, D. A. and Bennett, D. A. (2004). Diabetes mellitus and risk of Alzheimer disease and decline in cognitive function. *Archives of Neurology*, 61, 661–666.

Bassett, S. S., Chase, G. A., Folstein, M. and Regier, D. A. (1998). Disability and psychiatric disorders in an urban community: measurement, prevalence and outcomes.*Psychological Medicine*, 28, 509–517.

Brandt, J., Spencer, M. and Folstein, M. F. (1988). The Telephone Interview for Cognitive Status. *Neuropsychiatry, Neuropsychology and Behavioural Neurology*, 1, 111–117.

Burback, D., Molnar, F. J., St. John, P. and Man-Son-Hing, M. (1999). Key methodological features of randomized controlled trials of Alzheimer's disease therapy. Minimal clinically important difference, sample size and trial duration. *Dementia and Geriatric Cognitive Disorders*, 10, 534–540.

Cercy, S. P. and Bylsma, F. W. (1997). Lewy bodies and progressive dementia: a critical review and meta-analysis. *Journal of the International Neuropsychology Society*, 3, 179–194.

Crum, R. M., Anthony, J. C., Bassett, S. S. and Folstein, M. F. (1993). Population-based norms for the Mini-Mental State Examination by age and educational level. *JAMA*, 269, 2386–2391.

Cummings, J. L. (1997). The Neuropsychiatric Inventory: assessing psychopathology in dementia patients. *Neurology*, 48 (Suppl. 6), S10–16.

Devanand, D. P. *et al.* (1996). Depressed mood and the incidence of Alzheimer's disease in the elderly living in the community. *Archives of General Psychiatry*, 53, 175–182.

Evans, D. A. *et al.* (1997). Apolipoprotein E ϵ4 and incidence of Alzheimer disease in a community population of older persons. *JAMA*, 277, 822–824.

Flaten, T. P. (2001). Aluminum as a risk factor in Alzheimer's disease, with emphasis on drinking water. *Brain Research Bulletin*, 55, 187–196.

Folstein, M. F., Folstein, S. E. and McHugh, P. R. (1975). "Mini-mental State": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189–198.

Folstein, M. F. and Whitehouse, P. J. (1983). Cognitive impairment of Alzheimer disease. *Journal of Neurobehavior Toxicology and Teratology*, 5, 631–634.

Frank, R. M. and Byrne, G. J. (2000). The clinical utility of the Hopkins Verbal Learning Test as a screening test for mild dementia. *International Journal of Geriatric Psychiatry*, 15, 317–324.

Gatz, M. *et al.* (2006). Role of genes and environments for explaining Alzheimer disease. *Archives of General Psychiatry*, 63, 168–174

Han, L., Cole, M., Bellavance, J., McCusker, J. and Primeau, F. (2000). Tracking cognitive decline in Alzheimer's disease using the Mini-mental State Examination: a meta-analysis. *International Psychogeriatrics*, 12, 231–247.

Hardy, J. and Gwinn-Hardy, K. (1998). Genetic classification of primary neurodegenerative disease. *Science*, 282, 1075–1079.

Hatanpaa, K. J. *et al.* (2004). Most cases of dementia with hippocampal sclerosis may represent frontotemporal dementia. *Neurology*, 63, 538–542.

Hayden, K. M. *et al.* (2006). Vascular risk factors for incident Alzheimer disease and vascular dementia: the Cache County study. *Alzheimer Disease and Associated Disorders*, 20, 93–100.

Janofsky, J. S., McCarthy, R. J. and Folstein, M. F. (1992): The Hopkins competency assessment test: a brief method for evaluating patients' capacity to give informed consent. *Hospital and Community Psychiatry*, 43, 132–136.

Jicha, G. A. *et al.* (2006). Neuropathologic outcome of mild cognitive impairment following progression to clinical dementia. *Archives of Neurology*, 63, 674–681.

https://doi.org/10.1017/S1041610207005169 Published online by Cambridge University Press

Karweit, M., Brock, I. P. and Folstein, M. F. (1994). A digital-video medical record system for assessing demented patients. In G. Racagni, N. Brunell and S. Z. Langer (eds.), *Recent Advances in the Treatment of Neurodegenerative Disorders and Cognitive Symptoms, Vol. 7* (pp 31–38). Basel: S. Karger Press.

Lye, T. C. and Shores, E. A. (2000). Traumatic brain injury as a risk factor for Alzheimer's disease: a review. *Neuropsychological Review*, 10, 115–129.

Mohs, R. C. and Cohen, L. (1988). Alzheimer's Disease Assessment Scale (ADAS). *Psychopharmacology Bulletin*, 24, 627–628.

Morris, J. C. (1993). The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology*, 43, 2412–2414.

Nott, P. N. and Fleminger, J. J. (1975). Presenile dementia: the difficulties of early diagnosis. *Acta Psychiatrica Scandinavica*, 51, 210–217.

Pericak-Vance, M. A. *et al.* (1997). Complete genomic screen in late-onset familial Alzheimer disease: evidence for a new locus on chromosome 12. *JAMA*, 278, 1237–1241.

Qiu, W. Q. and Folstein, M. F. (2006). Insulin, insulin-degrading enzyme and amyloid-beta peptide in Alzheimer's disease: review and hypothesis. *Neurobiology of Aging*, 27, 190–198.

Rasmusson, D. X., Bylsma, F. W. and Brandt, J. (1995). Stability of performance on the Hopkins Verbal Learning Test. *Archives of Clinical Neuropsychology*, 10, 21–26.

Reisberg, B. *et al.* (1996). Behavioral pathology in Alzheimer's disease (BEHAVE-AD) rating scale. *International Psychogeriatrics* (Suppl. 3), 301–308, discussion (351–354).

Royall, D. R., Mahurin, R. K. and Gray, K. F. (1992). Bedside assessment of executive cognitive impairment: the executive interview. *Journal of the American Geriatrics Society*, 40, 1221–1226.

Samuel, W., Terry, R D., DeTeresa, N., Butters, N. and Masliah, E. (1994). Clinical correlates of cortical and nucleus basalis pathology in Alzheimer dementia. *Archives of Neurology*, 51, 772–778.

Stern, Y., Mayeux, R., Sano, M., Hauser, W. A. and Bush, T. (1987). Predictors of disease course in patients with probable Alzheimer's disease. *Neurology*, 37, 1649–1653.

Terry, R. D. *et al.* (1991). Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment. *Annals of Neurology*, 30, 572–580.

Tune, L. J. *et al.* (1991). Physostigmine in Alzheimer's disease: effects on cognitive functioning, cerebral glucose metabolism analyzed by positron emission tomography and cerebral blood flow analyzed by single photon emission tomography. *Acta Psychiatrica Scandinavica* (Suppl.), 366, 61–65.

Valenzuela, M. J. and Sachdev, P. (2006). Brain reserve and cognitive decline: a non-parametric systematic review. *Psychological Medicine*, 36, 1065–1073.

https://doi.org/10.1017/S1041610207005169 Published online by Cambridge University Press