Do MCI criteria in drug trials accurately identify subjects with predementia Alzheimer's disease?

P J Visser, P Scheltens, F R J Verhey

J Neurol Neurosurg Psychiatry 2005;76:1348–1354. doi: 10.1136/jnnp.2004.047720

See end of article for authors' affiliations

Correspondence to:
Dr Pieter Jelle Visser,
Department of Psychiatry
and Neuropsychology,
University of Maastricht,
PO Box 616, 6200 MD
Maastricht, Netherlands;
pj.visser@np.unimaas.nl

Received 4 July 2004
In revised form
29 October 2004
Accepted 30 January 2005

Subjects with mild cognitive impairment (MCI) often suffer from Alzheimer's disease in the predementia stage.¹ For this reason, people with MCI have been enrolled in trials with drugs that were effective in patients with Alzheimer-type dementia. However, previous studies have shown that not all MCI subjects have predementia Alzheimer's disease.^{2,3} In view of this well recognised heterogeneity of MCI, it is important to determine the extent to which drug trials have included subjects with MCI who indeed had predementia Alzheimer's disease. Our aim in the present study was to investigate whether MCI criteria used in previous and ongoing drug trials could accurately diagnose subjects with predementia Alzheimer's disease. We applied these criteria to a well characterised cohort of MCI patients whose cognitive outcome after five years of follow up was known and in whom the presence of predementia Alzheimer's disease at baseline could be inferred. We also investigated how exclusion criteria for depression and vascular co-morbidity affected the diagnostic accuracy. MCI trials applied exclusion criteria for depression or vascular comorbidity because these conditions may cause MCI. However, depressive symptomatology and vascular comorbidity often occur in subjects with predementia Alzheimer's disease, and exclusion of these may also decrease the diagnostic accuracy. Finally, because the prevalence of predementia Alzheimer's disease increases with age we investigated whether the diagnostic accuracy was dependent on age.

METHODS

Subjects

The diagnostic accuracy of MCI definitions was investigated retrospectively in a cohort of newly referred consecutive non-demented patients from the Maastricht Memory Clinic with

Background: Drugs effective in Alzheimer-type dementia have been tested in subjects with mild cognitive impairment (MCI) because these are supposed to have Alzheimer's disease in the predementia stage.

Objectives: To investigate whether MCI criteria used in these drug trials can accurately diagnose subjects with predementia Alzheimer's disease.

Methods: MCI criteria of the Gal-Int 11 study, InDDEx study, ADCS memory impairment study, ampakine CX 516 study, piracetam study, and Merck rofecoxib study were applied retrospectively in a cohort of 150 non-demented subjects from a memory clinic. Forty two had progressed to Alzheimer type dementia during a five year follow up period and were considered to have predementia Alzheimer's disease at baseline. Outcome measures were the odds ratio, sensitivity, specificity, and positive and negative predictive value.

Results: The odds ratio of the MCI criteria for predementia Alzheimer's disease varied between 0.84 and 11. Sensitivity varied between 0.46 and 0.83 and positive predictive value between 0.43 and 0.76. None of the criteria combined a high sensitivity with a high positive predictive value. Exclusion criteria for depression led to an increase in positive predictive value and specificity at the cost of sensitivity. In subjects older than 65 years the positive predictive value was higher than in younger subjects.

Conclusions: The diagnostic accuracy of MCI criteria used in trials for predementia Alzheimer's disease is low to moderate. Their use may lead to inclusion of many patients who do not have predementia Alzheimer's disease or to exclusion of many who do. Subjects with moderately severe depression should not be excluded from trials in order not to reduce the sensitivity.

no apparent cause for their cognitive impairment such as cerebrovascular disorders, brain trauma, or severe psychiatric disorders.⁴ Subjects were referred to the memory clinic by general practitioners (70%), psychiatrists (16%), neurologists (6%), or other physicians (8%). Subjects were reinvestigated two, five, and 10 years after the baseline visit. For the present study we selected all subjects older than 50 years who had been eligible for the five year follow up assessment ($n = 178$). They were considered to have predementia Alzheimer's disease at baseline if they met criteria for Alzheimer-type dementia during the five year follow up period. Outcome with respect to cognitive functioning was available for 150 subjects (82.3%). Of these, 42 (28%) had predementia Alzheimer's disease at baseline. The group of subjects without predementia Alzheimer's disease ($n = 108$) included one with frontotemporal dementia, one with dementia from other causes, three with vascular cognitive impairment, and one with Parkinson's disease at follow up. Cognitive outcome was not available for 28 subjects because they had died ($n = 5$), could not be traced ($n = 12$), had refused follow up ($n = 9$), or were not contacted for other reasons ($n = 2$). Subjects without cognitive outcome were older than those with cognitive outcome (68.3 years *v* 61.8 years), while sex, years of education, and scores on the global deterioration

Abbreviations: BDRS, Blessed dementia rating scale; CDR, clinical dementia rating scale; DSM-IV, *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition; Gal-Int-11, galantamine international 11 study; GDS, global deterioration scale; HDRS, Hamilton depression rating scale; HIS, Hachinski ischaemic scale; InDDEx, investigation into delay to diagnosis of Alzheimer's disease with Exelon study; LM, logical memory test; MCI, mild cognitive impairment; MMSE, mini-mental state examination; NINCDS-ADRDA, National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association; RAVLT, Rey auditory verbal learning task

Table 1 Eligibility criteria, mild cognitive impairment inclusion criteria, and exclusion criteria for depression and vascular comorbidity

Study	Eligibility criteria*			MCI inclusion criteria		Exclusion criteria	
	Age (years)	MMSE score	Other	Functional impairment	Test impairment	Depression	Vascular
Gal-Int-11	>50	–	–	CDR 0.5†	NYU <11†	Moderately severe depression	–
InDDEx	55 to 85	–	–	CDR 0.5†	NYU <9†	HDRS >12; HDRS item 1 >1; DSM-IV major depression	m-HIS >4
ADCS-MIS	55 to 90	>23	–	CDR 0.5†	Edu <8: LM-A <3† Edu 8–15: LM-A <5† Edu >15: LM-A <9†	HDRS >12	m-HIS >4
Ampakine	55 to 80	>24	–	CDR 0.5†	Edu <8: LM-A <5† Edu 8–15: LM-A <8† Edu >15: LM-A <13†	GDS >6†	m-HIS >4
Piracetam	50 to 90	–	Score of at least one point above the lowest possible score on 7 of 8 tests from a cognitive battery†	CDR 0.5†	LM immediate recall <10 or difference immediate and delayed recall LM >5†	HDRS >17†	–
Rofecoxib	>65	>23	BDRS: score <4 on sum items 1–12 and score on first 8 items all <1	CDR 0.5†	RAVLT summed immediate recall < 38	HDRS >12	m-HIS >4

*All studies required that the subject was not demented at baseline.

†Scale or test was not used in the Maastricht Memory Clinic. An equivalent scale or test was used if possible as described in the text.

BDRS, Blessed dementia rating scale; CDR, clinical dementia rating scale; Edu, years of education; GDS, geriatric depression scale; HDRS, Hamilton depression rating scale; LM-A, delayed recall score of story A from the logical memory test of the Wechsler memory scale; LM, logical memory test from the Wechsler memory scale-revised; m-HIS, modified Hachinski ischaemic scale; MMSE, mini-mental state examination; NYU, New York University paragraph recall.

scale (GDS),⁵ mini-mental state examination (MMSE),⁶ Hamilton depression rating scale (HDRS),⁷ and the delayed recall of the Rey auditory verbal learning task (RAVLT) did not differ.

Subjects gave informed consent before inclusion in the study. The study was approved by the medical ethics committee of the University Hospital Maastricht, Netherlands.

Clinical assessment and clinical diagnosis at baseline and follow up

At baseline subjects underwent a standardised assessment which included a history provided by the patient and a significant other; a psychiatric, neurological, and physical examination; clinical rating scales (MMSE,⁶ GDS,⁵ HDRS-17 items,⁷ the Blessed dementia rating scale (BDRS) part I,⁸ and the Hachinski ischaemic scale (HIS⁹); appropriate laboratory tests; a neuropsychological assessment; and neuroimaging as described elsewhere.¹⁰ The diagnosis of dementia and Alzheimer's disease was made according to the DSM-IV and NINCDS-ADRDA criteria.^{11–12} The follow up assessment consisted of a standardised questionnaire about medical history and cognitive complaints, the MMSE, the GDS, the HDRS, the BDRS, and a neuropsychological test protocol.^{13–14} If a subject declined to come for the follow up assessment, a telephone interview was conducted. This included a standardised questionnaire about medical history and cognitive complaints, and the Telephone Interview for Cognitive Status.¹⁵ The diagnosis of dementia and Alzheimer's disease at follow up was made by a neuropsychiatrist and a neuropsychologist who were unaware of the results of the baseline assessment and who made their diagnosis independently of each other. If there was disagreement about the clinical diagnosis, a consensus meeting was held and if no agreement was reached the subject was considered not demented.

Definition of MCI according to criteria from trials

MCI trials were selected from the review by Petersen¹ and included the galantamine international 11 (Gal-Int-11) study, the investigation into delay to diagnosis of Alzheimer's disease with Exelon (InDDEx) study,¹⁶ the Alzheimer's disease cooperative study–memory impairment study (ADCS-MIS),¹⁷ the ampakine CX 516 study,¹⁸ the piracetam study, and the Merck rofecoxib MCI study.^{19–20} Table 1 shows the eligibility criteria, inclusion criteria for MCI, and exclusion criteria for depression and vascular comorbidity of these studies. The criteria were applied retrospectively in the subjects from the Maastricht Memory Clinic. If scales and tests used in the trials had not been used in the memory clinic we chose equivalent scales or tests if possible, as described below.

Modification of eligibility criteria

We did not apply the requirement in the piracetam study that subjects should score at least one point above the lowest possible score on seven of eight tests from a cognitive battery as we did not use this battery or an equivalent battery.

Modification of MCI inclusion criteria

Instead of a score of 0.5 on the clinical dementia rating scale, we used a score of 3 on the GDS or a score of 2 on the GDS together with a score of at least 0.5 on the first eight items of the BDRS as a measure of functional impairment. Instead of the New York University (NYU) paragraph test used in the Gal-Int-11 and InDDEx studies and the logical memory (LM) test of the Wechsler memory scale used in the ADCS-MIS, ampakine, and piracetam studies we used a Dutch story recall test which consists of a 20 item story.²¹ So that we could use a similar cut off for the Dutch story recall test, we estimated the centile score relative to healthy control subjects to which the cut off scores on the NYU test and LM test corresponded and used this score as the cut off on the Dutch story recall

Table 2 Number of subjects meeting eligibility criteria and inclusion and exclusion criteria for mild cognitive impairment

	Subjects meeting eligibility criteria (n)	Subjects meeting MCI inclusion criteria (n)	Subjects meeting exclusion criteria (n)*		Subjects meeting MCI criteria and exclusion criteria (n)
			Depression	Vascular comorbidity	
Gal-Int 11	138	90	13 (2)	–	77
InDDEx	98	57	24 (12)	2 (1)	32
ADCS-MIS	97	41	16 (6)	0	25
Ampakine	90	48	14 (4)	2 (1)	33
Piracetam	138	76	15 (3)	–	61
Rofecoxib	41	34	9 (7)	0	25

*Numbers of subjects with predementia Alzheimer's disease given in brackets.

ADCS-MIS, Alzheimer's disease cooperative study-memory impairment study; Gal-Int-11, galantamine international study 11; InDDEx, investigation into delay to diagnosis of Alzheimer's disease with Exelon; MCI, mild cognitive impairment.

test. Centile scores of the NYU paragraph test were based on control subjects from the NYU (Kluger A, personal communication). The cut off score used in the Gal-Int-11 study corresponded to a score below the 53rd centile, and the cut off score used in the InDDEx study to a score below the 31st centile. As no normative data have been published for story A of the LM test used in the ADCS-MIS and ampakine studies, we multiplied the cut off by 2 and calculated centile scores based on published age norms for story A and story B together.²²

In the ADCS-MIS trial, the centile scores were dependent on age and education and varied between the 38th and 62nd centile for subjects with more than 15 years of education, between the 4th and 24th centile for subjects with 8–15 years of education, and between the 1st and 10th centile for subjects with less than 8 years of education. In the ampakine study, the centile scores were also dependent on age and education and varied between the 76th and 90th centile for subjects with more than 15 years of education, between the 24th and 50th centile for subjects with 8–15 years of education, and between the 4th and 24th centile for subjects with less than 8 years of education. The cut off on the immediate recall of the LM test used in the piracetam study corresponded with a centile score between the 1st and 9th centile depending on the age of the subject.²³ The cut off on the difference between the immediate recall and delayed recall of the LM test used in the piracetam trial corresponded approximately to a score below the 50th centile.²³

Modification of exclusion criteria

Instead of a score >6 on the geriatric depression rating scale we used a score >12 on the HDRS as an exclusion criterion in the ampakine study. Instead of a score >17 on the HDRS-21 items scale used in the piracetam study, we used a score >17 on the HDRS-17 items scale. The exclusion criterion of moderately severe depression used in the Gal-Int-11 study was operationalised as a diagnosis of major depression according to DSM-IV criteria together with a score on the HDRS >17.

Missing data

Of the 150 subjects with known cognitive outcome, several had missing data for one or more tests or rating scales. The HIS was not available for 27 subjects. As these subjects did not display major vascular pathology on clinical examination none of these subjects was excluded. The Dutch story recall test was not administered to 12 subjects. Subjects with missing data for this test were not included in the eligible sample for MCI criteria that made use of this test. Compared with subjects in whom the Dutch story recall test was administered, those without this test scored higher on the HDRS (14.4 v 10.2). Age, sex, educational level, MMSE score, delayed recall score on the RAVLT, or prevalence of predementia Alzheimer's disease did not differ. The number of subjects who were excluded from the eligible sample because of missing data for the Dutch story recall test was 12 for the Gal-Int-11 study and piracetam studies, 11 for the

Table 3 Baseline characteristics of subjects according to different definitions of mild cognitive impairment

	MCI definition in Gal-Int-11 study (n = 77)	MCI definition in InDDEx study (n = 32)	MCI definition in ADCS-MIS study (n = 25)	MCI definition in ampakine study (n = 33)	MCI definition in piracetam study (n = 61)	MCI definition in rofecoxib study (n = 25)
Age	64.2 (8.4)	66.6 (6.9)	68.2 (6.5)	67.8 (7.0)	64.5 (8.9)	70.9 (4.2)
Male:female (% male)	47:30 (61)	20:12 (63)	17:8 (68)	21:12 (64)	35:26 (57)	18:7 (72)
Years of education	10.2 (3.1)	10.8 (3.4)	11.1 (3.7)	11.0 (3.4)	10.4 (3.3)	10.7 (3.2)
Education (%)*						
Low	10 (13)	3 (9)	3 (12)	3 (9)	9 (15)	3 (12)
Middle	60 (78)	24 (75)	17 (68)	24 (73)	46 (75)	19 (76)
High	7 (9)	5 (16)	5 (20)	6 (18)	6 (10)	3 (12)
GDS score	2.7 (0.47)	2.6 (0.50)	2.7 (0.46)	2.7 (0.48)	2.7 (0.44)	2.8 (0.41)
GDS 2 (%)	25 (33)	14 (44)	7 (28)	11 (33)	16 (26)	5 (20)
GDS 3 (%)	52 (67)	18 (56)	18 (72)	22 (67)	45 (74)	20 (80)
HDRS score	8.6 (5.1)	4.7 (3.2)	5.2 (3.7)	5.6 (3.6)	8.6 (5.1)	7.1 (3.5)
m-HIS	1.4 (1.7)	1.0 (1.3)	0.96 (1.3)	1.0 (1.2)	1.3 (1.7)	0.84 (1.1)
MMSE score	27.6 (1.8)	27.1 (1.7)	26.7 (1.2)	27.3 (1.4)	27.3 (1.8)	27.4 (1.4)
Delayed recall AVLT (z score†)	–1.5 (1.2)	–1.7 (1.4)	–2.2 (1.1)	–2.0 (1.2)	–1.7 (1.2)	–1.5 (1.1)
Fluency animals (z score†)	–0.78 (0.83)	–0.96 (0.82)	–1.1 (0.84)	–0.99 (0.82)	–0.85 (0.79)	–0.91 (0.80)
Subjects with predementia AD (%)	33 (43)	18 (56)	19 (76)	23 (70)	29 (48)	13 (52)

Values are mean (SD) unless specified otherwise.

*Low: less than 8 years of education; middle: 8 to 15 years of education; high: 16 or more years of education.

†z score indicates the number of standard deviations from the average of a healthy control population.¹⁴ A negative z score indicates a score below the average. AD, Alzheimer's disease; ADCS-MIS, Alzheimer's disease cooperative study-memory impairment study; AVLT, auditory verbal learning task Gal-Int-11, galantamine international study 11; GDS, global deterioration scale; HDRS, Hamilton depression rating scale; InDDEx, investigation into delay to diagnosis of Alzheimer's disease with Exelon; MCI, mild cognitive impairment; m-HIS, modified Hachinski ischaemic scale; MMSE, mini-mental state examination.

Table 4 Diagnostic accuracy of definitions of mild cognitive impairment for predementia Alzheimer's disease

MCI definition*	Subjects in group A/B/C/D (n)†	OR	Sensitivity	Specificity	PPV	NPV
<i>MCI definition of the Gal-Int-11 study</i>						
Inclusion criteria only	43/55/5/35	5.4 (2.0 to 15)	0.88 (0.77 to 0.98)	0.44 (0.34 to 0.54)	0.39 (0.29 to 0.49)	0.90 (0.81 to 0.98)
Inclusion and exclusion criteria	54/44/7/33	5.8 (2.3 to 14)	0.83 (0.71 to 0.94)	0.55 (0.45 to 0.65)	0.43 (0.32 to 0.54)	0.89 (0.81 to 0.97)
<i>MCI definition of InDDEx study</i>						
Inclusion criteria only	32/27/9/30	4.0 (1.6 to 9.8)	0.77 (0.64 to 0.90)	0.54 (0.42 to 0.67)	0.53 (0.40 to 0.66)	0.78 (0.65 to 0.91)
Inclusion and exclusion criteria	45/14/21/18	2.8 (1.2 to 6.6)	0.46 (0.31 to 0.62)	0.76 (0.65 to 0.87)	0.56 (0.39 to 0.73)	0.68 (0.57 to 0.79)
<i>MCI definition of ADCS-MIS study</i>						
Inclusion criteria only	43/16/13/25	5.2 (2.1 to 12)	0.66 (0.51 to 0.81)	0.73 (0.62 to 0.84)	0.61 (0.46 to 0.76)	0.77 (0.66 to 0.88)
Inclusion and exclusion criteria	53/6/19/19	8.8 (3.1 to 25)	0.50 (0.34 to 0.66)	0.90 (0.82 to 0.98)	0.76 (0.59 to 0.93)	0.74 (0.63 to 0.84)
<i>MCI definition of ampakine study</i>						
Inclusion criteria only	36/21/5/28	9.6 (3.2 to 29)	0.85 (0.73 to 0.97)	0.63 (0.51 to 0.76)	0.57 (0.43 to 0.71)	0.88 (0.78 to 0.98)
Inclusion and exclusion criteria	47/10/10/23	11 (3.9 to 30)	0.70 (0.54 to 0.85)	0.82 (0.73 to 0.92)	0.70 (0.54 to 0.85)	0.82 (0.73 to 0.92)
<i>MCI definition of piracetam study</i>						
Inclusion criteria	54/44/10/32	4.9 (2.1 to 11.7)	0.80 (0.68 to 0.92)	0.55 (0.45 to 0.65)	0.42 (0.31 to 0.53)	0.87 (0.79 to 0.95)
Inclusion and exclusion criteria	66/32/11/29	5.4 (2.4 to 12)	0.73 (0.59 to 0.86)	0.67 (0.58 to 0.77)	0.48 (0.35 to 0.60)	0.86 (0.78 to 0.94)
<i>MCI definition of rofecoxib study</i>						
Inclusion criteria	5/14/2/20	3.6 (0.60 to 21)	0.91 (0.79 to 1.0)	0.26 (0.07 to 0.46)	0.59 (0.42 to 0.75)	0.71 (0.38 to 1.0)
Inclusion and exclusion criteria	7/12/9/13	0.84 (0.23 to 3.0)	0.59 (0.39 to 0.80)	0.37 (0.15 to 0.59)	0.52 (0.32 to 0.74)	0.44 (0.19 to 0.68)

Values are means (95% confidence intervals) unless stated.

*Inclusion criteria refer to criteria for mild cognitive impairment. Exclusion criteria refer to exclusion criteria for depression and vascular comorbidity.

†Group A, subjects eligible for the study who did not meet the criteria of the MCI definition and who had no predementia Alzheimer's disease; group B, subjects eligible for the study who met criteria of the MCI definition and who had no predementia Alzheimer's disease; group C, subjects eligible for the study who did not meet criteria of the MCI definition and who had predementia Alzheimer's disease; group D, subjects eligible for the study who met criteria of the MCI definition and who had predementia Alzheimer's disease.

ADCS-MIS, Alzheimer's disease cooperative study-memory impairment study; Gal-Int-11, galantamine international study 11; InDDEx, investigation into delay to diagnosis of Alzheimer's disease with Exelon; MCI, mild cognitive impairment; NPV, negative predictive value; OR, odds ratio; PPV, positive predictive value.

Table 5 Effect of age on diagnostic accuracy of mild cognitive impairment definitions for predementia Alzheimer's disease

MCI definition	Subjects in group A/B/C/D (n)*	OR	Sensitivity	Specificity	PPV	NPV
<i>MCI definition of Gal-Int-11 study</i>						
Age 50 to 64 years	47/33/3/6	2.8 (0.66 to 12)	0.67 (0.36 to 0.97)	0.59 (0.48 to 0.70)	0.15 (0.04 to 0.27)	0.94 (0.87 to 1.0)
Age 65 to 85 years	7/11/4/27	4.3 (1.0 to 18)	0.87 (0.75 to 0.99)	0.39 (0.16 to 0.61)	0.71 (0.57 to 0.85)	0.64 (0.35 to 0.92)
<i>MCI definition of InDDEx study</i>						
Age 55 to 64 years	32/8/3/5	6.9 (1.4 to 35)	0.63 (0.29 to 0.96)	0.80 (0.68 to 0.93)	0.38 (0.12 to 0.65)	0.92 (0.83 to 1.0)
Age 65 to 85 years	12/6/18/13	1.4 (0.43 to 4.9)	0.42 (0.25 to 0.59)	0.67 (0.45 to 0.88)	0.68 (0.48 to 0.89)	0.40 (0.22 to 0.58)
<i>MCI definition of ADCS-MIS study</i>						
Age 55 to 64 years	39/2/3/5	33 (4.3 to 244)	0.63 (0.29 to 0.96)	0.95 (0.89 to 1.0)	0.71 (0.38 to 1.0)	0.93 (0.85 to 1.0)
Age 65 to 85 years	14/4/16/14	3.1 (0.82 to 11)	0.47 (0.29 to 0.65)	0.78 (0.59 to 0.97)	0.78 (0.59 to 0.97)	0.47 (0.29 to 0.65)
<i>MCI definition of ampakine study</i>						
Age 55 to 64 years	35/5/3/5	12 (2.1 to 65)	0.63 (0.29 to 0.96)	0.88 (0.77 to 0.98)	0.50 (0.19 to 0.81)	0.92 (0.84 to 1.0)
Age 65 to 85 years	12/5/7/18	6.2 (1.6 to 24)	0.72 (0.54 to 0.90)	0.71 (0.49 to 0.92)	0.78 (0.61 to 0.95)	0.63 (0.41 to 0.85)
<i>MCI definition of piracetam study</i>						
Age 55 to 64 years	55/24/4/6	4.7 (1.1 to 20)	0.67 (0.36 to 0.97)	0.70 (0.60 to 0.80)	0.20 (0.06 to 0.34)	0.95 (0.89 to 1.0)
Age 65 to 85 years	10/8/8/23	3.6 (1.1 to 12)	0.74 (0.59 to 0.90)	0.56 (0.33 to 0.79)	0.74 (0.59 to 0.90)	0.56 (0.33 to 0.79)

Values are means (95% confidence intervals) unless stated.

*Group A, subjects eligible for the study who did not meet the criteria of the MCI definition and who had no predementia Alzheimer's disease; group B, subjects eligible for the study who met criteria of the MCI definition and who had no predementia Alzheimer's disease; group C, subjects eligible for the study who did not meet criteria of the MCI definition and who had predementia Alzheimer's disease; group D, subjects eligible for the study who met criteria of the MCI definition and who had predementia Alzheimer's disease.

ADCS-MIS, Alzheimer's disease cooperative study-memory impairment study; Gal-Int-11, galantamine international study 11; InDDEx, investigation into delay to diagnosis of Alzheimer's disease with Exelon; MCI, mild cognitive impairment; NPV, negative predictive value; OR, odds ratio; PPV, positive predictive value.

InDDEx study, nine for the ADCS-MIS study, and eight for the ampakine study. The BDRS was not scored in one subject and this individual was excluded from the eligible sample of the rofecoxib study.

Statistical analysis

For each MCI definition we first determined the sample that would be eligible for the study. This sample consisted of all subjects in the age range of the study who did not meet the criteria for dementia at baseline. Next we applied the inclusion criteria for MCI. Then we applied the exclusion criteria for depression or vascular comorbidity. Three sets of analyses were performed. First, we calculated the diagnostic accuracy of the MCI definition using both inclusion criteria and exclusion criteria in the subjects who were eligible for the study. Second, we calculated the diagnostic accuracy of the MCI definition using only inclusion criteria. Third, we calculated the diagnostic accuracy using both inclusion criteria and exclusion criteria in a young subgroup (age 50 or 55 to 64) and an old subgroup (age 65 or above). For each analysis we defined four groups: subjects eligible for the study without predementia Alzheimer's disease who did not meet the criteria of the MCI definition (group A); subjects eligible for the study without predementia Alzheimer's disease who met the criteria of the MCI definition (group B); subjects eligible for the study with predementia Alzheimer's disease who did not meet the criteria of the MCI definition (group C); and subjects eligible for the study with predementia Alzheimer's disease who met the criteria of the MCI definition (group D). Outcome measures were the odds ratio $((A \cdot D)/(B \cdot C))$, sensitivity $(D/(C+D))$, specificity $(A/(A+B))$, positive predictive value $(D/(B+D))$, and negative predictive value $(A/(A+C))$. Data are presented as means and 95% confidence intervals (CI).

RESULTS

Diagnostic accuracy of MCI definitions

Table 2 shows the number of subjects who were eligible for the study, the number of eligible subjects who met the inclusion criteria for MCI, the number of subjects with MCI who were excluded because of depression or vascular comorbidity, and the number who met both the MCI inclusion and exclusion criteria. The baseline characteristics of subjects meeting inclusion and exclusion criteria is shown in table 3. The odds ratio, sensitivity, specificity, positive predictive value, and negative predictive for predementia Alzheimer's disease of each of the MCI definitions is shown in table 4.

Effect of exclusion criteria for depression and vascular comorbidity on diagnostic accuracy

Table 4 shows that after application of the exclusion criteria the odds ratios of the criteria in the Gal-Int-11, ADCS-MIS, and piracetam studies increased and those of the criteria in the InDDEx, ampakine, and rofecoxib studies decreased. The sensitivity and negative predictive value decreased and the specificity increased for all MCI criteria. The positive predictive value of all criteria except those in the rofecoxib study increased.

Effect of age on diagnostic accuracy

The effect of age on diagnostic accuracy of the criteria in the rofecoxib study was not tested as that study included only subjects older than 65 years. Table 5 shows the effect of age on the diagnostic accuracy. The positive predictive value was higher and the specificity and negative predictive value were lower in the older than in the younger subgroup for all criteria. For most criteria, the odds ratio was lower and the sensitivity higher in the older than in the younger subgroup.

DISCUSSION

MCI criteria used in recent drug trials have a low to moderate diagnostic accuracy for predementia Alzheimer's disease. Application of exclusion criteria for depression and vascular comorbidity led to an increase in the positive predictive value and specificity at the cost of the sensitivity. The diagnostic accuracy of the MCI criteria for predementia Alzheimer's disease was dependent on age.

Diagnostic accuracy of MCI criteria for predementia Alzheimer's disease

There was marked heterogeneity in diagnostic accuracy for predementia Alzheimer's disease between the different MCI criteria. The highest odds ratios were seen for MCI criteria from the ampakine study ($OR = 11$) and ADCS-MIS study ($OR = 8.8$). Nevertheless, these odds ratios are below the observed minimum for a good diagnostic test (25). The odds ratios of the other MCI criteria were much lower (between 0.84 and 5.8). The variability in diagnostic accuracy may depend on the variability in the definition of cognitive impairment. Although all studies used the same definition of functional impairment (a score of 0.5 on the CDR), there were marked differences for the definition of impairment on memory tests. Memory impairment was defined as a score below the 53rd centile in the Gal-Int-11 study, below the 31st centile in the InDDEx study, below the 9th centile in the ADCS-MIS study (this is the average centile score given the age and education distribution in our sample), below the 31st centile in the ampakine study (this is also the average percentile score given the age and education distribution in our sample), below the 50th centile in the piracetam study, and below the 46th centile in the rofecoxib study (the centile score was estimated from data of healthy subjects older than 65 years from the Maastricht aging study).²¹ As can be seen in table 2, studies with a strict cut off score for memory impairment had a higher positive predictive value than those with a lenient cut off score. It is remarkable that the InDDEx and Gal-Int-11 studies used a permissive definition of memory impairment on the NYU paragraph test, as it has been shown that the best score on this test to identify subjects with predementia Alzheimer's disease among those with MCI is a score below 6, which is well below the cut off scores used in InDDEx and Gal-Int-11 studies.²⁴

Effect of exclusion criteria for depression and vascular comorbidity on diagnostic accuracy

The application of exclusion criteria for depression and vascular comorbidity led to an increase in the odds ratio for predementia Alzheimer's disease (indicating an improvement in overall diagnostic accuracy) for the MCI criteria of the Gal-Int-11, ADCS-MIS, ampakine, and piracetam studies, and a decrease in odds ratio for the MCI criteria of the InDDEx and rofecoxib studies. The specificity and positive predictive value typically increased, while the sensitivity and negative predictive value decreased. These effects mainly reflected the exclusion criteria for depression because 88–100% of the excluded subjects were excluded because of depression. The small number of exclusions because of vascular comorbidity is probably because we had already excluded at baseline those subjects in whom the cognitive impairment was thought to be due to vascular lesions. The decrease in odds ratio of the MCI criteria of the InDDEx and rofecoxib studies was probably the result of the relatively large number of excluded subjects who had predementia Alzheimer's disease. The effect of the exclusion criteria on the sensitivity correlated well with the severity of the threshold for depression: the fall in sensitivity was low for the MCI criteria of the Gal-Int-11 and piracetam studies which used an high threshold for depression and high for the MCI criteria of the

InDDEX study which used a low threshold. The large effect of depression cut off on sensitivity is consistent with previous studies that showed a high prevalence of mild depressive disorders in subjects with predementia Alzheimer's disease.^{14 25 26}

Effect of age on diagnostic accuracy

The positive predictive value was higher in the older than in the younger age group, which is probably because the prevalence of predementia Alzheimer's disease was higher in the older age group. The odds ratio was higher in the younger than in the older subgroup for the MCI criteria of the InDDEX, ADCS-MIS, ampakine, and piracetam studies. One possible explanation for this observation is that in the younger subgroup the exclusion criteria for depression more often excluded subjects without than with predementia Alzheimer's disease, while in the older age group the exclusion criteria for depression more often excluded subjects with than without predementia Alzheimer's disease.

Limitations

A limitation of the study is that we defined predementia Alzheimer's disease as conversion to Alzheimer's disease-type dementia after a five year follow up period, because MCI subjects may convert to Alzheimer's disease-type dementia at longer follow up intervals. Thus we may have made a false negative diagnosis of predementia Alzheimer's disease in some cases, which may have led to an underestimation of the positive predictive value and to an overestimation of the negative predictive value. However, the number of misclassified subjects is likely to be small as preliminary data from the 10 year follow up of a subset of patients indicate that only 10% of the non-demented subjects at the five year follow up converted to dementia during the next five years.

A second limitation is the relatively small sample size for several of the MCI definitions and age subgroups. This reduced statistical power and resulted in large 95% confidence intervals.

Another limitation is that we could not always use the same tests and rating scales that were used in the trials, which may limit the comparability of our results with that of the trials. Baseline characteristics of subjects from the ADCS-MIS and InDDEX trials suggests that nevertheless we have included a similar population, because subjects included in the ADCS-MIS trial had similar scores on the MMSE, GDS, delayed recall, and verbal fluency as subjects who met criteria of the ADCS-MIS study in our sample¹⁷; and subjects included in the InDDEX trial had similar scores on the MMSE and GDS as subjects who met the criteria of the InDDEX study in our sample (table 1).¹⁶ In addition, the finding that the positive predictive value (which is the same as the conversion rate) was low for subjects meeting the MCI criteria of the InDDEX and Rofecoxib studies is consistent with the fact that the follow up of these studies had to be extended because of a low conversion rate.¹

Another difference between the present study and the trials is that we selected subjects from a clinical setting at the time of the baseline assessment, while in the trials subjects could also have been recruited at follow up assessments and from other settings (for example, they could have been recruited by advertisements). This difference in subject selection may have affected the diagnostic accuracy because of referral and follow up bias. Nevertheless, the way we selected subjects is representative of the setting in which drugs would be prescribed.

Finally, our memory clinic is located at a psychiatry outpatient clinic and the effect of the exclusion criteria for depression on diagnostic accuracy may be less in settings in

which the prevalence of depression or the awareness of the diagnosis is lower.

Implications for interpretation of present and ongoing trials

The findings may have important implications for the interpretation of present and ongoing trials. The overall diagnostic accuracy was low to moderate, which means that many patients will be treated who do not have predementia Alzheimer's disease, or that many patients with predementia Alzheimer's disease will not be included. It seems useful to plan subanalyses in subjects older than 65 years as these are more likely to have predementia Alzheimer's disease. For studies that used a permissive definition of memory impairment, it seems recommendable to plan analysis in a subsample of subjects with more severe memory impairment.

Implications for design of future trials

As the use of MCI criteria alone will not lead to an accurate selection of subjects with predementia Alzheimer's disease, a major challenge is to find other ways to identify such individuals. An alternative approach would be to combine a number of markers of predementia Alzheimer's disease such as age, the MMSE score, degree of functional impairment, impairment on neuropsychological tests, medial temporal lobe atrophy, and the apolipoprotein E-e4 allele.⁴ Another challenge is to exclude subjects with depression related cognitive impairment without excluding depressed subjects with predementia Alzheimer's disease. One option would be to exclude only subjects with severe depression, as people with predementia Alzheimer's disease typically have mild to moderate depression and score below 21 on the HDRS.¹⁴ Another possibility is to exclude only young depressed subjects, as people with depression related cognitive impairment are younger than those with predementia Alzheimer's disease.¹⁴

NOTE ADDED IN PROOF

After submission of the final version we obtained normative data for story A of the logical memory (LM) test from the Wechsler memory scale from the Chicago health and aging project (Bennett D, personal communication). As these normative data may provide better estimates of the centile scores than those we used in the paper we have recalculated the diagnostic accuracy of the MCI definitions of the ADCS-MIS and ampakine studies using the new data. The cut off points for story A of the LM test used in the ADCS-MIS study now corresponded with an age and education corrected centile score between the 17th and 56th centile (on average, the 23th centile), and the cut off points used in the ampakine study with a centile score between the 39th and 63rd centile (on average, the 45th centile). This indicates that according to the new normative data the severity of the memory impairment was less than with the normative data used in the main text. Thirty two subjects now met the inclusion and exclusion criteria of the ADCS-MIS study and 44 met the inclusion and exclusion criteria of the ampakine study. The odds ratio of the ADCS-MIS MCI definition for predementia Alzheimer's disease was 7.5, the sensitivity 0.58, the specificity 0.83, the positive predictive value 0.69, and the negative predictive value 0.75. The odds ratio of the ampakine MCI definition for predementia Alzheimer's disease was 3.9, the sensitivity 0.70, the specificity 0.63, the positive predictive value 0.52, and the negative predictive value 0.78. The effect of the exclusion criteria on the diagnostic accuracy of the MCI definition of the ADCS-MIS study was similar to the effect described in table 4. For the MCI definition of the ampakine study, however, the odds ratio decreased after application of the exclusion criteria,

from 6.8 to 3.9. The odds ratio of the ADCS-MIS MCI definition for predementia Alzheimer's disease was 12 for subjects younger than 65 and 3.4 for those older than 65. The odds ratio of the ampakine MCI definition for predementia Alzheimer's disease was 4.4 for subjects younger than 65 and 1.8 for subjects older than 65.

In conclusion, with the new normative data, the estimates of the specificity, positive predictive accuracy, and overall diagnostic accuracy were lower than the estimates based on the normative data that were used in the main text. This effect was largest for the MCI definition of the ampakine study. The use of the new normative data would not change the conclusions of our paper.

ACKNOWLEDGEMENTS

We would like to thank Paula Hemdal (UCB group) for providing additional information on the MCI definition used in the piracetam study and Cristopher Lines (Merck Research Laboratories) for providing additional information regarding the MCI definition used in the Merck rofecoxib MCI study.

Authors' affiliations

P J Visser, F R J Verhey, Department of Psychiatry, University Hospital Maastricht and Alzheimer Centre Limburg, Maastricht, Netherlands

P Scheltens, Department of Neurology, Alzheimer Centre, VU Medisch Centrum, Amsterdam, Netherlands

Competing interests: PS is an associate editor of JNNP but had no role in the reviewing or acceptance of this paper.

REFERENCES

- Petersen RC. Mild cognitive impairment clinical trials. *Nat Rev Drug Discov* 2003;**2**:646–53.
- Bennett DA, Wilson RS, Schneider JA, et al. Natural history of mild cognitive impairment in older persons. *Neurology* 2002;**59**:198–205.
- Grober E, Lipton R, Hall C, et al. Memory impairment on free and cued selective reminding predicts dementia. *Neurology* 2000;**54**:827–32.
- Visser PJ, Verhey FRJ, Scheltens P, et al. Diagnostic accuracy of the Preclinical AD Scale (PAS) in cognitively mildly impaired subjects. *J Neurol* 2002;**249**:312–19.
- Reisberg B, Ferris SH, De Leon MJ, et al. The global deterioration scale for assessment of primary degenerative dementia. *Am J Psychiatry* 1982;**139**:1136–9.
- Folstein MF, Folstein SE, McHugh PR. "Mini-mental state", a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;**12**:189–98.
- Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*, 1960;**23**:56–62.
- Blessed G, Tomlinson BE, Roth M. The association between quantitative measures of dementia and of senile changes in the cerebral grey matter of elderly subjects. *Br J Psychiatry* 1968;**114**:797–811.
- Hachinski VC, Iliff LD, Zilhka E, et al. Cerebral blood flow in dementia. *Arch Neurol* 1975;**32**:632–7.
- Verhey FRJ, Jolles J, Ponds RWHM, et al. Diagnosing dementia: a comparison between a monodisciplinary and multidisciplinary approach. *J Neuropsychiatry Clin Neurosci* 1993;**5**:78–85.
- American Psychiatric Association A. *Diagnostic and statistical manual of mental disorders*, 4th ed. Washington DC: American Psychiatric Association, 1994.
- McKhann G, Drachman D, Folstein M, et al. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Working Group under the auspices of the Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology* 1984;**34**:939–44.
- Visser PJ, Verhey FRJ, Ponds RWHM, et al. Course of objective memory impairment in non-demented subjects attending a memory clinic and predictors of outcome. *Int J Geriatr Psychiatry* 2000;**15**:363–72.
- Visser PJ, Verhey FRJ, Ponds RWHM, et al. Distinction between preclinical dementia and depression. *J Am Geriatr Soc* 2000;**48**:479–84.
- Brandt J, Spencer M, Folstein M. The Telephone Interview for Cognitive Status. *Neuropsychiat Neuropsychol Behav Neurol* 1988;**1**:111–17.
- Feldman H, Scheltens P, Scarpini E, et al. Behavioral symptoms in mild cognitive impairment. *Neurology* 2004;**62**:1199–201.
- Grundman M, Petersen RC, Ferris SH, et al. Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Arch Neurol* 2004;**61**:59–66.
- Johnson SA, Simmon VF. Randomized, double-blind, placebo-controlled international clinical trial of the Ampakine CX516 in elderly participants with mild cognitive impairment: a progress report. *J Mol Neurosci* 2002;**19**:197–200.
- Lines CR, McCarroll KA, Lipton RB, et al. Telephone screening for amnesic mild cognitive impairment. *Neurology* 2003;**60**:261–6.
- Thal L, Ferris S, Kirby L, et al. A randomized, double-blind, study of rofecoxib in patients with mild cognitive impairment. *Neuropsychopharmacology*, 2005, March 02 [Epub ahead of print].
- In: Jolles J, Houx P, van Bostel MPJ, Ponds RWHM, eds. *Maastricht aging study: determinants of cognitive aging*. Maastricht: Neuropsych Publishers, 1995.
- Wechsler D. *Wechsler memory scale, 3rd ed, Administration and scoring manual*. San Antonio: The Psychological Corporation, 1997.
- Ivnik RJ, Malec JF, Smith GE, et al. Mayo's Older Americans Normative Studies: WMS-R norms for ages 56 to 94. *Clin Neuropsychologist* 1992;**6**(suppl 1):48–81.
- Kluger A, Ferris SH, Golomb J, et al. Neuropsychological prediction of decline to dementia in nondemented elderly. *J Geriatr Psychiatry Neurol* 1999;**12**:168–79.
- Tierney MC, Boyle E, Lam RE, et al. Do depressive symptoms in memory-impaired elders predict probable Alzheimer's disease? *Aging Mental Health* 1999;**3**:88–93.
- Berger AK, Fratiglioni L, Forsell Y, et al. The occurrence of depressive symptoms in the preclinical phase of AD: a population-based study. *Neurology* 1999;**53**:1998–2002.