

# Literature Review of the Clock Drawing Test as a Tool for Cognitive Screening

Elisabete Pinto Ruth Peters

Care of the Elderly, Faculty of Medicine, Imperial College London, London, UK

## Key Words

Clock Drawing Test • Cognitive screening • Mild cognitive impairment

## Abstract

**Background/Aims:** The Clock Drawing Test (CDT) is used in clinical practice for the screening of cognitive disorders. This systematic review aims to present and discuss the CDT scoring methods available in the literature, in order to find differences in administration and utility of the CDT. **Methods:** A literature search was carried out in Medline (1966 to June 2008), Psycinfo (1967 to June 2008) and EMBASE (1980 to June 2008). **Results:** All studies showed good interrater and test-retest reliabilities. The correlation with other standard screening tests was statistically significant in most studies, but the results were influenced by age, education and language. In studies that included patients with mild or questionable dementia, the CDT had a low sensitivity and variable specificity. **Conclusion:** The CDT has the characteristics of a good screening method for moderate and severe dementia. However, the scoring method used and potential confounders need to be taken into consideration.

Copyright © 2009 S. Karger AG, Basel

## Introduction

Research into the early screening of cognitive disorders, specifically cognitive impairment and dementia, has increased over the last two decades. One of the driving forces behind this is the extended life span of the ageing world population with the consequent increase in the incidence of dementia and mild cognitive impairment (MCI) [1].

MCI has been defined as a transitory state between normal cognition and dementia [2, 3]. However, more recently, Petersen and O'Brien [4] described MCI as an entity distinct from normal ageing and from Alzheimer's disease (AD). There are various subtypes of MCI characterised in the literature. For example, MCI can be characterised by an amnesiac syndrome, impairment in a single non-memory domain of cognition or slight cognitive decline in more than one domain [5]. Approximately 10% of the patients with amnesiac MCI will develop into AD type dementia each year [3] and early intervention in this group could improve the treatment of AD. AD is the most common form of dementia and presents with memory loss. However, disturbances with executive cognitive functioning may precede the memory decline [6]. These result in difficulties with activities of daily living [7]. There are new potential pharmacological treatments which may have the capability of slowing down the pro-

**Table 1.** Inclusion and exclusion criteria for selection of studies

Inclusion criteria	Exclusion criteria
CDT to assess cognitive function	Raters not blinded to patient diagnosis
Well-defined control group	Additional psychiatric disorder
Standard criteria for diagnosis of dementia	

gression of dementia, although further research is required. To harvest the benefits of any possible health interventions, early screening tests are essential. Due to these reasons, interest in the early screening of MCI and dementia has increased [8].

A study on the prevalence of dementia using the Short Portable Mental Status Questionnaire of approx. 4,000 patients aged 60 or over estimated that only 25% of people with AD or MCI are recognised in general practice [9]. Comprehensive neuropsychological batteries of tests such as the Blessed Test of Orientation, Concentration and Memory [10] are very useful and perform well in identifying MCI and dementia. However, these battery tests are lengthy and difficult to use as routine in the primary care setting. Therefore, cognitive screening instruments such as the Mini-Mental State Examination (MMSE) [11] are preferable, although the MMSE still takes 20–30 min to administer and can miss patients in the early stages of cognitive impairment [8]. Although the MMSE somewhat reflects visuoconstructive deficits and executive dysfunction, it is mostly recognised as a verbal test [12, 13]. The level of education, gender, age, ethnicity and social class of the patients [14, 15] also frequently affect these screening tests.

Although dementia is a major indirect cause of hospitalisation and use of health care resources in the elderly [16], there is a reluctance of physicians to systematically assess cognition. Bush et al. [17] conducted a survey of general practitioners and found that although 82% of general practitioners believed that dementia should be screened for, only 24% routinely screened their patients. Their main reasons were lack of time, fear of offending the patients and inadequacy of easily administered tests.

Cognitive screening tools that are easy to administer and have high sensitivity and specificity are essential. There is also the requirement for instruments that are valid across cultures and easy to administer in a variety of settings. The Clock Drawing Test (CDT) has been used

for decades as a screening test and is used within diverse languages and backgrounds [18–20]. The CDT taps into a series of cognitive domains that may be impaired early in AD, such as: verbal understanding, memory, spatial knowledge, abstract thinking, planning, concentration and visuoconstructive skills [21, 22].

Despite the various advantages of the CDT, including its simplicity, speed of administration in a busy practice and the potential to be less offensive to patients, there are still many important aspects that require further study. These issues include: the most appropriate scoring system to be used, the training required by the rater (naïve vs. professional) and at what level the test should be performed (general practitioner vs. specialised service) [23].

A literature review search showed that the most recent CDT review paper available on screening was for the year 2000 [22]. Since 2000, further interest in the CDT as a cognitive screening tool has originated studies with specific populations and also several studies comparing scoring systems. An up-to-date systematic review of the CDT will bring together these recent developments. It will also allow us to identify areas where further study is required.

## Methods

A computer search was performed on the following databases: Psychinfo (1967 to June 2008), Medline (1966 to June 2008) and EMBASE (1980 to June 2008). The following keywords were used: 'Clock Drawing Test', 'Clock Drawing Task', 'Clock Test', 'Clock Drawing' or 'CDT'. 'Clock' was also used with 'cognitive decline', with 'cognitive impairment' or with 'executive function'. The searches were restricted to participants  $\geq 65$  years of age and in English, French, Portuguese, Spanish or Italian language. A total of 305 abstracts were identified for further selection. Seventy-six research papers in which the CDT was used as a screen for cognitive function were selected and read by both authors independently. A total of 41 research papers were finally selected for this review in order to include only those studies which met the following selection criteria (table 1): use of the CDT to assess cognitive function; inclusion of a well-defined control group, and use of standard criteria for the diagnosis of dementia, that is, either the Diagnostic and Statistical Manual of the American Psychiatric Association (DSM-III/IV-TR) or the National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA/NINCDS-AIREN) criteria.

Studies were excluded when the raters scoring the CDT were not blinded to the patient diagnosis as this could lead to a potential bias in scoring. Studies in which participants were suffering from any additional psychiatric disorder that may have affected performance on cognitive testing were also excluded.

**Table 2.** Characteristics of clock drawing scoring methods

Reference	Test	Pre-drawn circle	Time setting	Scoring method and score range	Original cut-off for abnormal
Shulman et al. [23, 24]	drawing	yes	11:10	5 points modified to 6 hierarchical point scale (1–5, 0–5)	>2 (1986) <3 (1993)
Wolf-Klein et al. [25]	drawing	yes	no	10 hierarchical patterns (1–10)	<7
Sunderland et al. [26]	drawing	no	2:45	10 hierarchical point scale (1–10)	<6
Mendez et al. [27] (CDIS)	drawing	no	11:10	20-item scale based on errors each scored 0/1 (0–20)	<19
Rouleau et al. [28]	drawing and copying	no	11:10	10-point scale and 6-item qualitative scale (0–10)	≤7
Tuokko et al. [29]	drawing, clock setting, clock reading	yes	11:10	25 defined errors in 7 error categories (0–no ceiling)	>2
Watson et al. [30]	drawing	yes	no	clock divided into quadrants: errors in the first to third quadrant = 1 point; error in the forth quadrant = 4 points (0–7)	>3
Manos and Wu [31]	drawing	yes	11:10	clock divided into eighths, points given for numbers and hands in right place (0–10)	<8
Cahn et al. [32]	drawing	no	11:10	quantitative scale derived from Rouleau et al. [28] and also 8 qualitative error types (0–10 points, 1–8 errors)	quantitative ≤7 qualitative ≥1
Lam et al. [33]	drawing, reading and time setting	yes	3:00	10-item scale (0–10)	>3
Borson et al. [34] (CERAD)	drawing	no	8:20	4 anchored points (0–3)	>0
Royall et al. [35] (CLOX)	drawing and copying	no	1:45	15-point scale (0–15)	<10
Cacho et al. [36]	drawing and copying	no	11:10	10 points (system modified from Rouleau et al. [28] and Sunderland et al. [26])	≤6
Jitapunkul et al. [37] (CCSS)	drawing	yes	11:10	clock divided in quadrants, 5 items marked 0–2 points (0–10)	<7
Lin et al. [38]	drawing and copying	yes	10:10	clock divided in quadrants, 16 items marked 0/1 (0–16); short version only 3 items (0–3)	<11 short version <3
Heinik et al. [39] (CDT-MIA)	drawing	no	11:10	20 items (0–33)	<23

## Results

The CDT was used in the 41 studies selected for this literature review as a tool for cognitive screening. This test can be used in clinic, in both the primary and secondary settings. It is quick to administer and to score and it has shown good correlation with other commonly used tests such as the MMSE.

However, there are considerable limitations, probably the most important being the variation of results achieved by different researchers using variant methods for its scoring. The studies also used patients with various backgrounds, medical history and degrees of cognitive dysfunction. Consequently, many studies using the same scoring methods found considerable difference in the sensitivity and specificity of the CDT. Consideration is also necessary for variables that can affect the validity of the test such as the patients' language, culture and level of education. All this may reduce the robustness of the test and its usefulness as a screening tool.

In what follows, we include sections for the administration, scoring methods, studies comparing scoring methods, influence of language and education and the screening of questionable dementia. We present the studies included in this review in the chronological order in which they were published.

### *Administration of the CDT*

The various authors and the different scoring systems use slightly different instructions and methodology for administering the CDT to the patients. The most common instructions are: 'Please draw a clock face, placing all the numbers on it. Now set the time to 10 past 11.' However, certain differences in administration should be considered, particularly the use of a predrawn circle, time setting and additional copying or time reading commands (table 2).

Some tests include an already predrawn circle such as seen in Manos and Wu [31] and Tuokko et al. [29]. Other authors prefer to observe the patients drawing their own

clock circumference as that can indicate some degree of impairment. The disadvantage is that if the patient starts off with a poorly drawn circle, at times merely due to age-related problems such as tremor, the remainder of the test may be compromised.

Most authors ask for a time setting with the exception of Wolf-Klein et al. [25] and Watson et al. [30]. Frequently the time asked is 10 past 11. This allows for the detection of problems in concrete thinking, specially the tendency to pull the minute hand towards the number 10. Some authors specifically ask to place the hands and numbers, whereas others administer the test using a very simple instruction of drawing a clock with a set time.

The inclusion of copying and time setting or reading tests in addition to clock drawing by some authors such as Lam et al. [33] could improve the test's predictive validity, but also increases its complexity and time of administration, reducing one of the main good features of the CDT, its speed of completion. The correlation between patient cognitive status and clock copying was found to be less significant than that with clock drawing [33]. Cacho et al. [36] also used the copy command and saw that patients with AD showed an improvement with the copy command compared to the verbal command. The difference of improvement between both commands improved the distinction between healthy controls and AD patients. The verbal command seems to imply a greater participation of the memory aspects, altered in the AD patients, whereas the copy command uses visual-constructive components that may be normal in these patients. Memory may be affected at the early stages of AD, whereas visual-constructive components may be intact in the early stages of the disease.

#### *CDT Scoring Systems*

Table 3 shows the properties of the most common scoring methods found in the literature and includes studies by the authors who developed these scoring systems and also later studies. Several measures are presented as they were reported, including sensitivity and specificity as well as interrater (agreement between raters) and test-retest (agreement between repeated tests) reliabilities.

One of the oldest scoring systems used was by Shulman et al. in 1986 [23]. They developed a 5-point scale of hierarchical errors with the higher score for a failed attempt to draw a clock. In a subsequent study, this scale was reversed and the maximum points given to a perfect clock [24]. Beinhoff et al. [40] used this system in triage

to determine its usefulness in patients with various degrees of dementia. The CDT seemed useful only for the detection of AD, but not of MCI.

Sunderland et al. [26] used a 10-point anchored system based on preset criteria with an arbitrary cut-off at 6 points. They found that interrater reliability was high in clinicians and non-clinicians. The CDT was significantly correlated to other measures of dementia severity, such as the Dementia Rating Scale, but not to age or duration of symptoms in either AD patients or controls [26]. Kirby et al. [41] used the same scoring system, but in a more heterogeneous control group of community-dwelling participants without depression or dementia, and obtained reasonable percentages of sensitivity (76%) and specificity (81%) [41].

Wolf-Klein et al. [25] tested patients who were admitted consecutively to a nursing facility without preselection, although the group with AD was older than the normal group. The system uses 10 clock patterns identified from a previous pilot study and a score less than 7 indicates abnormality.

The system by Tuokko et al. [29] includes clock setting and clock reading in addition to clock drawing. The scoring for the clock drawing uses a series of 25 error types with no ceiling for the total score where more than 2 errors indicate impairment. The errors with highest differences between patients and controls were omissions and misplacement of numbers. O'Rourke et al. [42] used the Tuokko system for the detection of early cognitive impairment. The patients were referrals to AD clinics accessed at two different times. The authors found no differences in the CDT scores with gender or level of education, although there was a considerable age difference between the groups and they obtained lower sensitivity and specificity than the original study by Tuokko et al. They found that substitution errors contributed mostly towards the test discrimination [42].

The Clock Drawing Interpretation Scale (CDIS) by Mendez et al. [27] uses 20 points distributed between general impression, placement of numbers and hands with a score of higher than 18 as being normal. The authors found that the presence of the number 2 and the correct location of the minute hand were the items most frequently absent and were absent in all AD patients [27]. The CDIS system was also used by Lee et al. [43] alongside the Sunderland scoring system. They found that the correlation with the CAMCOG test (the cognitive and self-contained part of the Cambridge Examination for Mental Disorders of the Elderly) was significant for the Sunderland method, but not for the Mendez system. The speci-

**Table 3.** Original CDT studies and later studies

Method	Reference	Number (cases/controls)	Sensitivity %	Specificity %	Interrater reliability	Test-retest reliability	Correlation with other screening tests (r)
Shulman	Shulman et al. [23, 24]	75	86	72	0.75		MMSE = -0.65; SPMSQ = -0.66; GDS = -0.32
	Beinhoff et al. [40]	183 232 (114/57)	cut-off 1-86 cut-off 2-71	cut-off 1-60 cut-off 2-95	0.96		
Sunderland	Sunderland et al. [26]	150 (67/83)			0.86		GDS r = 0.56; DRS r = 0.59; BDRS r = 0.51; SPMSQ r = 0.59, p < 0.001
	Kirby et al. [41]	648 (41/523)	76	81			
Wolf-Klein	Wolf-Klein et al. [25]	312 (182/130)	75	94			
Tuokko	Tuokko et al. [29]	120 (58/62)	92	86	0.90-0.95	0.73	
	O'Rourke et al. [42]	59-22 new cases	cut-off <5-84	cut-off <5-73			
CDIS	Mendez et al. [27]	72 (46/26)			0.94	0.70-0.78	Rey Complex Figure Test r = 0.66, p < 0.001; Symbol Digit Modalities Test r = 0.65, p < 0.001 CAMCOG not significant Sunderland - CAMCOG praxis r = 0.71, p < 0.001
	Lee et al. [43] (also Sunderland)	60 (30/30)	73 67	77 97			
Watson	Watson et al. [30]	76 (40/36)	87	82	0.90 0.93	0.82	
	Casartelli et al. [44]	232 (25/207)	80	63			
Manos and Wu	Manos and Wu [31]	106 (63/43)	76	78	0.88-0.97	0.94	Trail Making Test Part A r = -0.48, p < 0.001; MMSE r = 0.50, p < 0.001; Block Design Test r = 0.56, p < 0.001
	Manos [45]	37 (14/23)	71	82			
Cahn	Cahn et al. [32]	353 (116/237)	88	63	0.95		
	Woodard et al. [8]	200 (39/161)	67	54			
Lam	Lam et al. [33]	106 (53/53)	83	79	0.97		MMSE r = -0.686, p < 0.01
	Chan et al. [46]	85 (51/34)	89	47			
Cacho	Cacho et al. [36]	112 (56/56)	93	94			CAMCOG memory subscale r = 0.7, p < 0.01
	Cacho et al. [47]	140 (70/70)					
CERAD	Borson et al. [34]	295 (170/125)	82	92	0.97	0.97	CASI r = -0.80, p < 0.001; MMSE r = -0.79, p < 0.001
	Scanlan and Borson [48]	249 (129/120)					
CLOX	Royal et al. [35]	85 (33/52)					EXIT25 r = -0.78, p < 0.001; MMSE r = 0.76, p < 0.001
CCSS	Jitapunkul et al. [37]	48 (12/36)	100	94			CMT r = 0.66, p < 0.001; WAIS r = 0.76, p < 0.01
Lin	Lin et al. [38]	403 (144/259)	0.67 short 0.73	d 0.75 short 0.66	0.99	0.9	CASI (Chinese version) r = 0.73, p < 0.01; MMSE r = 0.73, p < 0.01
CDT-MIA	Heinik et al. [39]	139 (93/46)	91	80	0.97		MMSE r = 0.50, p < 0.001 MMSE r = 0.77, p < 0.001; CAMCOG r = 0.77, p < 0.001
	Heinik et al. [49]	114 (88/26)					

SPMSQ = Short Portable Mental Status Questionnaire; GDS = Global Deterioration Scale; DRS = Dementia Rating Scale; BDRS = Blessed Dementia Rating Scale; CAMCOG = Cambridge Cognitive Examination Score; CASI = Cognitive Abilities Screening Instrument; EXIT25 = Executive Interview; CMT = Chula Mental Test; WAIS = Wechsler Adult Intelligence Scale.



ficity was also low for the Mendez system [43]. In the original Mendez study, a very high MMSE score of more than 28 was required for the control group originating a very high cut-off and consequently higher specificity [27].

The Clock Completion Test of Watson et al. [30] is an objective and simple scoring method. The subject is asked to place all the numbers in the clock, but not to set a time. Consequently, the scoring is only based on the position of the numbers in the clock face. No hands are required or scored and so some sensitivity is lost. The authors report that the number of digits in the 4th quadrant (9–12) had the best agreement with the diagnosis of dementia. To score the clock, this is divided into quadrants and a score of 1 point is given for any error in the first 3 quadrants and 4 points for any error in the 4th quadrant. A score of 0–3 is considered normal [30]. Casartelli et al. [44] also used the Clock Completion Test, but found a less good performance than in the original study. They suggested that this may be due to the weight placed on errors occurring in the 4th quadrant when asymmetry in number position was also found to be considerably high in the normal participants [44].

The Manos and Wu scoring system [31] is an objective scoring method which can be used by less trained staff. For the scoring, a transparent circle is divided into eighths and points 0–10 are awarded for numbers in the correct eighth and also for hands in the right place and right proportion. An arbitrary cut-off was set up at 7 points. They found that age affected the score in the control group [31]. A subsequent study [45] using the same system tried to identify very mild AD patients with an MMSE score >23 amongst consecutive ambulatory patients. The sensitivity was 71% compared with 76% for the original study, where patients had a mean MMSE score of 20.

The system by Cahn et al. [32] uses a double scale with quantitative points (0–10) and qualitative elements (1–8 errors). Woodard et al. [8] used the Cahn system within a battery of tests and found no difference in CDT scores between MCI and normal participants, but differences were seen between patients with neurocognitive disorder and normal participants. They did not find the CDT to be useful for a brief screening battery [8].

The system established by Lam et al. [33] uses scores up to 10 points with a score higher than 3 indicating impairment. Like the system developed by Cahn, this system also uses both quantitative and qualitative elements and in addition the functions of copying and time setting. Chan et al. [46] compared the CDT, scored using the Lam system, to the MMSE in a population from Hong

Kong. They found that there was a correlation between both tests although sensitivity and specificity were lower for the CDT than for the MMSE [46].

Cacho et al. [36] also used a method based on a 10-point scale with a cut-off of 6. They used both drawing and copying commands. In a later study, the same authors saw an improvement in patients with AD between the verbal and copy command, whereas no improvement was observed in the controls. This difference in improvement between both commands increased the CDT discrimination between the control and patient groups [47].

Borson et al. [34] used a CDT system within a battery, the MiniCog, comprising of the CDT and a simple memory test (three-word recall). The CDT was scored using a scoring system with 4 anchored points and a cut-off of 0. This system was also used in another study [48] where its results were compared with those of naïve raters classifying clocks simply as normal/abnormal. Comparison of naïve raters with expert raters using the CERAD system had an overall agreement of 89%, although it was considerably lower for mild impairment (60%).

Royall et al. [35] developed the CLOX test, a CDT scoring system, which they mention is specifically designed to measure executive control functions. The patient is asked to draw a clock on an empty page and later to copy a clock. The authors suggest that the difference between these tests can be a measure of executive control function. They compared the CLOX with the EXIT25, a measure of executive function, using two separate groups of outpatients and community volunteers. The differences between the groups were retained for the CDT after adjusting for gender, age and education. The CLOX was correlated with the EXIT25 and the MMSE [35].

Jitapunkul et al. [37] developed and used the Clock Completion Scoring System to score the CDT in Thai patients with memory complaints. The system uses a 10-point score with a cut-off at 7 points. They reported problems in administering the CDT to an illiterate population and the preference of using a predrawn circle.

Lin et al. [38] derived a simplified system to differentiate early to mild AD in a population in Taiwan. Age and education were significantly related to the CDT, although the correlation was low. They also developed a simplified CDT with only three items (hour hand, number 12 and difference between hands) and reported a good comparison between the simplified and the full version of the scoring system.

**Table 4.** Comparison studies

Reference	Number (case/control)	Scoring system	Sensitivity <sup>1</sup> , %	Specificity <sup>1</sup> , %	Correlation between scoring systems	ROC (AUC)
Brodaty and Moore [50]	56 (28/28)	Shulman Sunderland Wolf-Klein	86 57 (8/9–79) 36 (8/9–79)	96 100 (8/9–93) 100 (8/9–89)		from highest to lowest: Shulman, Sunderland, Wolf-Klein
Tuokko et al. [29]	493 (217/276)	Shulman Tuokko Wolf-Klein Watson	93 91 74 59	48 50 72 67		0.79 0.78 0.69 0.67
Storey et al. [51]	127 (72/55)	Shulman Mendez Sunderland Wolf-Klein Watson	93 96 69 43 69	55 26 58 86 44		0.79 0.78 0.71 0.71 0.65
Storey et al. [19]	93 (49/44)	Sunderland Mendez Shulman Wolf-Klein CERAD Watson	86 98 90 78 90 82	35 16 28 58 28 30		0.72 0.70 0.66 0.66 0.64 0.60
Scanlan et al. [52]	80 (60/20)	CERAD Shulman Sunderland Wolf-Klein Mendez Manos and Wu Lam	64 80 88 88 76 80 80	95 79 61 42 91 81 74	concordance with CERAD system – kappa coefficients 0.62, $p < 0.01$ 0.38, $p < 0.01$ 0.25, $p < 0.01$ 0.77, $p < 0.01$ 0.71, $p < 0.01$ 0.56, $p < 0.01$	
Schramm et al. [12]	123 (79/44)	Shulman Sunderland Wolf-Klein Watson Manos and Wu	81 56, (8/9) 85 39, (8/9) 81 56, (1/2) 77 67, (8/9) 81	79 91, (8/9) 80 95, (8/9) 73 80, (1/2) 50 86, (8/9) 75		from highest to lowest: Sunderland, Shulman, Manos and Wu, Wolf- Klein, Watson
Van der Burg et al. [53]	473 (387/86)	CERAD Shulman	97 96	32 42	0.77, $p < 0.001$	
Yamamoto et al. [54]	219 (130/89)	Sunderland Rouleau Cahn	43 57 75	95 93 76	Sunderland/Rouleau 0.84, $p < 0.001$ ; Sunderland/Cahn 0.86, $p < 0.001$ ; Rouleau/Cahn 0.98, $p < 0.001$	

<sup>1</sup> Figures in parentheses show results for cut-off values other than the one used in the original study.

### Comparison Studies of Scoring Methods

Comparisons between studies are difficult due to the diverse patient samples used. Further, the various reference standards and CDT scoring methods used by the authors make it difficult to arrive at definite conclusions.

The most recent studies have compared the results between several scoring methods (table 4) and mostly use the area under the receiver operating characteristics

(ROC) curve. The area under the ROC curve is perhaps a more unbiased method to determine the efficiency of a screening test as it shows the relationship between sensitivity and specificity.

Brodaty and Moore [50] compared the Shulman, Sunderland and Wolf-Klein methods. Shulman had the highest discriminatory power according to the ROC curve. Wolf-Klein showed the lowest power, and the authors suggested that this may be because the clock hands are

not scored in this method and so potentially relevant information is lost. Age was correlated with all scores when both groups were combined, but for the AD patient group the CDT scores were not correlated with age. Shulman's original cut-off point for AD was confirmed in this study, but for the Sunderland and Wolf-Klein methods the optimum cut-off points, that is those giving the best trade-off between sensitivity and specificity, were found to be different from the ones of the original authors. Brodaty and Moore excluded patients with severe dementia and so used a considerably different population from Sunderland or Wolf-Klein, and this could explain the difference in the optimum cut-off points [50].

Tuokko et al. [29] compared 4 methods and reported that the area under the ROC curve was lowest for the Watson and Wolf-Klein methods and highest for Shulman and Tuokko. The intrarater and interrater reliability was high for all methods, but significantly higher for the Tuokko and Watson methods [29].

Storey et al. [19, 51] published two research papers comparing various scoring methods in an Australian population of English and non-English speakers. They found that all methods had similar areas under the ROC curves, but were different in specificity and sensitivity. Although the Shulman and Mendez systems performed best, there was only a limited degree of accuracy for the detection of dementia in English-speaking patients [51] and even less in non-English speakers [19]. They concluded that the CDT was not adequate as a screening test for dementia in their sample.

Scanlan et al. [52] compared 7 scoring methods and in addition an intuitive system by naïve raters classifying clocks as normal or abnormal. The concordance between naïve raters (normal/abnormal) and the various scoring systems was high for all methods apart from Sunderland and Wolf-Klein. When level of dementia (Clinical Dementia Rating Scale, CDR) was analysed separately, the CERAD and Mendez methods were more sensitive at detecting mild-moderate dementia. The Wolf-Klein system performed poorly, failing to detect some cases of severe dementia. The authors found that naïve raters focused more on time setting than on skewed number placement [52].

Schramm et al. [12] tested whether the addition of the CDT to the MMSE or to the Short Performance Test (SKT) would improve screening. They compared 5 scoring methods and found that all methods correlated significantly with MMSE and SKT and had high interrater reliabilities. Shulman presented a reasonable sensitivity (81%) and specificity (79%) with its original cut-off point that could

not be improved by changing the cut-off point. However, changing the cut-off point for the other methods ameliorated the results (table 4). Using the area under the ROC curve, the highest discriminatory power was found for Sunderland followed by the Shulman system. The authors reported that a combination of the CDT with MMSE or the SKT leads to an increase in sensitivity [12].

Van der Burg et al. [53] analysed 2 systems, the simple CERAD scoring system and the Shulman method. Good correlation was found between both methods and no differences were found for sensitivity or specificity. The interrater reliability was better for the 4-item method and the authors concluded that a simple scoring system can be equally good.

Yamamoto et al. [54] compared 3 scoring systems determining sensitivity and specificity for all the possible cut-off points. The Cahn system seemed to have the highest discriminatory power. All methods were independent of years of education and depression and correlated with MMSE scores.

#### *Influence of Language and Level of Education*

The characteristics of the subject population such as language, cultural background and level of education can influence the usefulness and validity of the CDT test. Various studies have analysed the effect of several variables, especially the level of education of the patients. The results are contradictory, with several studies detecting a relation of the CDT score to education, whereas some did not find such a correlation.

Brodaty and Moore [50] found a correlation of the CDT score with years of education for the Shulman and Sunderland, but not the Wolf-Klein scoring system. Borson et al. [34] analysed the effect of the level of education and language spoken using patients from an Alzheimer's centre and controls from the general community. Demented patients were significantly less educated and older than non-demented patients. The CDT scores were lower for poorly educated participants, but did not differ with age. The differences between the groups were maintained after adjusting for education. Age had no effect on the scores of normal participants, but older demented patients had lower scores than younger ones. For non-English, less educated participants, the CDT gave the best balance between specificity and sensitivity [34].

Storey et al. [19, 51] published two studies in an Australian population of English- and non-English-speaking participants. They found that the specificity of the clock test was reduced when the population studied was of non-English-speaking background (table 4).



**Table 5.** Influence of education and language on the CDT scoring

Reference	Number (case/control)	Scoring method	Variable	Mean CDT score (case/control)	Correlation with variable
Brodsky and Moore [50]	56 (28/17)	Shulman Sunderland Wolf-Klein	education		Pearson's: $r = -0.35$ , $p < 0.01$ $r = 0.32$ , $p < 0.05$ $r = 0.2$ , NS
Borson et al. [34]	295 (170/125)	CERAD	education language (English vs. non-English)	$0.13 \pm 0.5$ vs. $2.1 \pm 1.2$	English sens spec Non-English sens spec $\leq 8$ years of education $\geq 9$ years of education
Leung et al. [55]	132 (66/66)	Lam	education	$5.8 \pm 3.16$ vs. $2.3 \pm 2.54$	Pearson's: AD patients $-0.25$ , $p < 0.05$ ; controls $-0.53$ , $p < 0.01$ ; minimum education $< 2$ years (cut-off 5/6): sens 73, spec 61, PPV 74 elementary education 2–6 years (cut-off 2/3): sens 75, spec 79, PPV 95 middle or higher education $> 6$ years (cut-off 0/1): sens 75, spec 67, PPV 89

sens = % Sensitivity; spec = % specificity; PPV = positive predictive value.

Leung et al. [55] analysed the effect of education on the CDT in AD patients (CDR-1) compared to community controls (table 5). The level of education was divided into three groups and although all scores correlated with education, the authors reported that the positive predictive value was higher for the medially educated group. The optimum cut-off points were, however, very different for the three groups and were much higher for patients with minimal education. The low cut-off point for the highest educated group suggests that the CDT is less useful for the screening of AD in this population. The CDT may be too simple for highly educated patients, even when some degree of cognitive impairment is already in place. The lower specificity found for the lowest educated group could be due to the use of the pen for the first time or infrequently (this is a potential problem for illiterate patients). The test was best for participants with elementary education [55].

However, a few studies found the CDT scores to be independent of the years of education [42, 54].

#### *Assessment of the CDT Validity in Screening Very Mild and Questionable Dementia*

The potential usefulness of the CDT is probably best applied when screening patients with very mild or questionable dementia. When the levels of dementia are severe or even moderate, symptoms become evident and a clinical diagnosis can be made more easily. In the literature, however, the populations used have more frequently than not mixed levels of dementia mostly at a moderate level. Below are some results of studies that specifically used patients with mild or questionable dementia (table 6).

Lee et al. [43] analysed the discrimination between mild and very mild AD patients, although only 9 patients fell in the latter category. They used both the Mendez and Sunderland scoring methods and found that the sensitivities for the very mild cases were poor due to a very wide standard deviation of the CDT scores. All patients progressed at least to mild dementia in the CDR Scale (CDR score = 1) at follow-up [43].

Powlishta et al. [56] compared CDT scores to the CDR levels and have shown that for individuals with very mild dementia (CDR score = 0.5), the CDT scores were not significantly different from normal individuals and the sensitivity obtained was very low.

Contrary to the two above studies, Esteban-Santillan et al. [18] analysed the use of the CDT for very mild AD with the Mendez and Rouleau scoring systems and found that the positive and negative predictive values were good in very mild AD. They selectively avoided questionable as well as moderate dementia, but the study was retrospective with records from an Alzheimer's institute and healthy research volunteers and the patients had a CDR score of 1, and not 0.5 as in the previous studies.

Seigerschmidt et al. [57] analysed the use of the CDT in questionable dementia. The CDR score of the patients used was 0.5, the same as in mild dementia. However, they used a mixed sample of patients and not only those who showed progression to overt dementia. Correlation of the CDT with other tests was low. There was a statistically significant difference in CDT scores between questionable dementia and normal participants and there were significant associations with education, age and gender. The CDT score decreased with age, and was low-

**Table 6.** Use of CDT for screening mild or questionable dementia

Reference	Number (case/ control)	Scoring method	Score (patients)	Mean CDT score				Sensitivity %				Specificity %	Correlation with other tests
				controls	CDR = 0.5	CDR = 1	CDR = 2	total	0.5	1	2		
Lee et al. [43]	60 (30/30)	Sunderland	CDR = 0.5 (9) CDR = 1 (17)	8.9 ± 1.4	6.3 ± 3	4.6 ± 2.1	2.2 ± 1.5	67	33	77	100	97	CAMCOG praxis r = 0.71, p < 0.001 NS
		Mendez	CDR = 2 (4)	19.1 ± 0.9	16.4 ± 4.7	13.4 ± 6	7 ± 8.1	73	44	82	100	77	
Esteban- Santillan et al. [18]	80 (41/39)	Rouleau	CDR = 1					93				42	
		Mendez	MMSE = 25 ± 1.5					46				79	
				controls	CDR = 0.5	CDR = 1	CDR = 2/3	total	0.5	≥1			
Powlishta et al. [56]	75 (60/15)	Manos and Wu		8.4 ± 2.4	6.9 ± 3.1	3.1 ± 3.2	1.5 ± 2.3	75	44	97		80	
		Mendez	CDR = 0.5 (25)	18.2 ± 2.3	16.3 ± 4	9.2 ± 6.9	5.4 ± 4.9	82	60	97		60	
		Rouleau	CDR = 1 (21)	8.0 ± 1.9	7 ± 2.3	3.3 ± 2.5	2.0 ± 1.5	58	20	86		93	
		Sunderland	CDR = 2/3(14)	8.4 ± 1.6	7.6 ± 2.4	4.1 ± 2.7	2.6 ± 1.1	57	20	83		93	
Seiger- schmidt et al. [57]	238 (99/139)	Manos and Wu	CDR = 0.5 MMSE = 25.5 ± 2.5	cases 7.1 ± 2.8 controls 8.4 ± 2.5				67				66	MMSE r = 0.32, p < 0.01 VF r = 0.32, p < 0.01 SKT r = 0.13, p < 0.05
				cases vs. controls				total	mild			total	mild
Connor et al. [58]	100 (50/50)	Wolf-Klein (<9)	mild dementia (DRS = 116–125)	7.3 ± 2.1 vs. 9.3 ± 0.8				78	60			78	75
		Rouleau (<8)		6.2 ± 2.4 vs. 8.8 ± 0.7				74	50			88	85
		Watson (>4)		4 ± 2.3 vs. 1.9 ± 1.8				52	40			84	80
				controls	CDR = 0.5	CDR = 1		QD	AD			QD	AD
Chiu et al. [59]	116 (76/40)	Rouleau	CDR = 0.5 (34) CDR = 1 (42)	9.1 ± 1.4	8.2 ± 1.5	6.95 ± 2.86		74	60			56	72

CDR: 0.5 – very mild/questionable dementia, 1 – mild, 2 – moderate, 3 – severe. CAMCOG = Cambridge Cognitive Examination Score; VF = Verbal Fluency Test; QD = questionable dementia.

er in women and higher in more educated participants. Overall, however, they found the CDT performance less good than the other tests, with a low sensitivity and specificity and to be influenced by confounders [57].

Connor et al. [58] compared several scoring methods in a sample of mildly impaired and moderately impaired patients. For the mildly impaired group, the sensitivity was at best 60%. The MMSE had the highest area under the ROC curve for the mild group and the addition of the CDT did not statistically increase the area under the ROC curve. The more cognitively impaired the patient group, the better the CDT is at differentiating, which decreases the usefulness of the CDT in mild cognitively impaired patients [58].

A recent study by Chiu et al. [59] compared the usefulness of the CDT in screening for questionable dementia with the CASI test and also found that the specificity was good for the detection of AD cases, but very low for questionable dementia cases.

## Discussion

In the literature, the CDT systems have shown to be lower in sensitivity and specificity when compared with the original studies by the authors who developed the systems. This may be due to different populations and control groups used. For example, studies conducted in spe-

cialist AD clinics such as the one by Tuokko et al. [29] would show higher sensitivities as patients have already frequently been referred to the clinic for memory problems. Other studies use selective control groups, excluding patients with memory problems or a family history of dementia [25], or a history of psychological or neurological illness [27]. The use of general community controls leads to a lower specificity, but it reflects the heterogeneous population a clinician encounters in practice [41].

It is generally observed that the sensitivity and specificity measures are adequate only when the decision criterion or cut-off point is agreed upon. Several studies using different populations found that the sensitivity and specificity could be improved by changing the cut-off point. The area under the ROC curve shows how the true-positive proportion (vertical axis) varies with the false-positive proportion (horizontal axis) as the decision criterion varies, and it is found to be a good measure of predictive accuracy. Several studies compared various scoring methods and although the results vary, systems based on a hierarchical scale, such as the Shulman and Sunderland, have higher areas under the ROC curve and so possibly higher discriminatory power. The Watson and Wolf-Klein systems did not perform as well in comparison studies. Neither of these two methods asks the patient to set a specific time when drawing the clock, and so the tests do not score the hands, leading to a loss of relevant information [12].

The correlation of the CDT with other screening tests, including the 'gold standard' MMSE, is good in most studies [31, 39]. Some authors argue that there may be a rationale for using both the MMSE and the CDT whilst screening for MCI or dementia, as the MMSE measures mostly verbal skills and so could miss patients with early dementia [60]. However, this would considerably increase the time of administration and so impair one of the characteristics of a good screening test.

Although professionals and specially trained scorers provide good interrater reliability, studies have shown that non-professional scorers have also provided good interrater reliabilities. For example, Sunderland et al. [26] showed that interrater reliabilities were as good for medical and health trained personnel as for non-clinicians. Similarly, Scanlan and Borson [48] used naïve raters and obtained good results with moderately demented patients, although results were not as good for mildly demented patients.

Variables such as language and education were also found to influence the performance of the CDT. Literacy could be a problem and could particularly affect the elderly participants in areas with a high illiteracy rate [37]. Leung et al. [55] showed that the CDT seems to perform better in participants with a medium level of education. Possibly, for participants with higher education, the CDT is too simple and in the early stages of cognitive decline these patients may retain a minimum level of ability to complete the CDT successfully.

All studies presented in this review showed that the CDT can be used to screen for moderate and severe dementia. However, studies that have used patients specifically with mild or questionable dementia have found that the sensitivity is considerably lower than for patients with moderate dementia. Although the CDT may be a useful test, further issues that may need to be considered include the fact that a screening test is most useful when it is directed at the early stages of the disease before any moderate or severe symptoms can be observed. More severely demented individuals would hopefully be recognised in practice without the need for screening [18].

In conclusion, the CDT is a useful screening test for the clinician, although its use in detecting early and mild cases of dementia is limited.

## References

- 1 Mashta O: Number of people in UK with dementia will more than double by 2050. *BMJ* 2007;334:447.
- 2 Flicker C, Ferris SH, Reisberg B: Mild cognitive impairment in the elderly: predictors of dementia. *Neurology* 1991;41:1006–1009.
- 3 Peterson RC, Smith GE, Waring SC: Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol* 1999;56:303–308.
- 4 Peterson RC, O'Brien J: Mild cognitive impairment should be considered for DSM-V. *J Geriatr Psychiatry Neurol* 2006;19:147–154.
- 5 Petersen RC, Doody R, Kurtz A: Current concepts in mild cognitive impairment. *Arch Neurol* 2001;58:1985–1992.
- 6 Binetti G, Magni E, Padovani A, Cappa SF, Bianchetti A, Trabucchi M: Executive dysfunction in early Alzheimer's disease. *J Neurol Neurosurg Psychiatry* 1996;60:91–93.
- 7 Royall DR, Cabello M, Polk MJ: Executive dyscontrol: an important factor affecting the level of care received by elderly retirees. *J Am Geriatr Soc* 1998;46:1519–1524.
- 8 Woodard JL, Dorsett ESW, Cooper JG, Hermann P, Sager MA: Development of a brief cognitive screen for mild cognitive impairment and neurocognitive disorder. *Aging Neuropsychol Cogn* 2005;12:299–315.

- 9 Callahan CM, Henrie HC, Tierney WM: Documentation and evaluation of cognitive impairment in elderly primary care patients. *Ann Intern Med* 1995;122:422–429.
- 10 Blessed G, Tomlinson BE, Roth M: The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects. *Br J Psychiatry* 1968;114:797–811.
- 11 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;12:189–198.
- 12 Schramm U, Berger G, Müller R, Kratzsch T, Peters J, Frölich L: Psychometric properties of Clock Drawing Test and MMSE or Short Performance Test (SKT) in dementia screening in a memory clinic population. *Int J Geriatr Psychiatry* 2002;17:254–260.
- 13 Royall DR, Mahurin RK: EXIT25, QED, and DSM-IV: very early Alzheimer's disease. *J Neuropsychiatry Clin Neurosci* 1994;6:62–64.
- 14 Teng EL, Hasegawa K, Homma A, Imai Y, Larson E, Graves A, Sugimoto K, Yamaguchi T, Sasaki H, Chiu D: The Cognitive Abilities Screening Instrument (CASI): a practical test for cross-cultural epidemiological studies of dementia. *Int Psychogeriatr* 1994;6:45–58.
- 15 Tombaugh TN, McIntyre NJ: The Mini-Mental State Examination: a comprehensive review. *J Am Geriatr Soc* 1992;40:922–935.
- 16 Narain P, Rubenstein LZ, Wieland GD, Rosbrook B, Strome LS, Pietruszka F, Morley JE: Predictors of immediate and 6-month outcomes in hospitalized elderly patients. The importance of functional status. *J Am Geriatr Soc* 1988;36:775–783.
- 17 Bush C, Kozak J, Elmslie T: Screening for cognitive impairment in the elderly. *Can Fam Physician* 1997;43:1763–1768.
- 18 Esteban-Santillan C, Praditsuwan R, Ueda H, Geldmacher DS: Clock Drawing Test in very mild Alzheimer's disease. *J Am Geriatr Soc* 1998;46:1266–1269.
- 19 Storey JE, Rowland JTJ, Basic D, Conforti DA: Accuracy of the Clock Drawing Test for detecting dementia in a multicultural sample of elderly Australian patients. *Int Psychogeriatr* 2002;14:259–271.
- 20 Agrell B, Dehlin O: The clock-drawing test. *Age Ageing* 1998;27:399–403.
- 21 Tuokko H, Hadjistavropoulos T, Miller A, Beattie BL: The clock test: a sensitive measure to differentiate normal elderly from those with Alzheimer disease. *J Am Geriatr Soc* 1992;40:579–584.
- 22 Shulman KI: Clock-drawing: is this the ideal cognitive screening test? *Int J Geriatr Psychiatry* 2000;15:548–561.
- 23 Shulman KI, Shedletsky R, Silver IL: The challenge of time: clock-drawing and cognitive function in the elderly. *Int J Geriatr Psychiatry* 1986;1:135–140.
- 24 Shulman KI, Gold DP, Cohen CA, Zuccherro CA: Clock-drawing and dementia in the community: a longitudinal study. *Int J Geriatr Psychiatry* 1993;8:487–496.
- 25 Wolf-Klein GP, Silverstone FA, Levy AP, Brod MS: Screening for Alzheimer's disease by clock drawing. *J Am Geriatr Soc* 1989;37:730–734.
- 26 Sunderland T, Hill JL, Mellow AM, Lawlor BA, Gundersheimer J, Newhouse PA, Grafman JH: Clock drawing in Alzheimer's disease, a novel measure of dementia severity. *J Am Geriatr Soc* 1989;37:725–729.
- 27 Mendez MF, Ala T, Underwood KL: Development of scoring criteria for the clock drawing task in Alzheimer's disease. *J Am Geriatr Soc* 1992;40:1095–1099.
- 28 Rouleau I, Salmon DP, Butters N, Kennedy C, McGuire K: Quantitative and qualitative analyses of clock drawings in Alzheimer's and Huntington's disease. *Brain Cogn* 1992;18:70–87.
- 29 Tuokko H, Hadjistavropoulos T, Rae S, O'Rourke N: A comparison of alternative approaches to the scoring of clock drawing. *Arch Clin Neuropsychol* 2000;15:137–148.
- 30 Watson YI, Arfken CL, Birge SJ: Clock completion: an objective screening test for dementia. *J Am Geriatr Soc* 1993;41:1235–1240.
- 31 Manos PJ, Wu R: The ten point clock test: a quick screen and grading method for cognitive impairment in medical and surgical patients. *Int J Psychiatry Med* 1994;24:229–244.
- 32 Cahn DA, Salmon DP, Monsch AU, Butters N, Wiederholt WC, Corey-Bloom J: Screening for dementia of the Alzheimer type in the community: the utility of the Clock Drawing Test. *Arch Clin Neuropsychol* 1996;11:529–539.
- 33 Lam LC, Chiu HF, Chan C, Chan WF, Li SW, Wong M: Clock-face drawing, reading and setting tests in the screening of dementia in Chinese elderly adults. *J Gerontol B Psychol Sci Soc Sci* 1998;53:353–357.
- 34 Borson S, Brush M, Gil E, Scanlan J, Vitalino P, Chen J, Cashman J, Sta Maria MM, Barnhart R, Roques J: The Clock Drawing Test: utility for dementia detection in multiethnic elders. *J Gerontol Med Sci* 1999;54A:M534–M540.
- 35 Royall DR, Mulroy AR, Chiodo LK, Polk MJ: Clock drawing is sensitive to executive control: a comparison of six methods. *J Gerontol Psychol Sci* 1999;54B:328–333.
- 36 Cacho J, Garcia-Garcia R, Arcaya J, Vicente JL, Lantada N: A proposal for the application of the Clock Drawing Test in Alzheimer's disease. *Rev Neurol* 1999;28:648–655.
- 37 Jitapunkul S, Worakul P, Kiatprakoth J: Validity of clinical use of the clock-drawing test in Thai elderly patients with memory problems. *J Med Assoc Thai* 2000;83:342–347.
- 38 Lin KN, Wang PN, Chen C, Chiu YH, Kuo CC, Chuang YY, Liu HC: The three-item clock-drawing test: a simplified screening test for Alzheimer's disease. *Eur Neurol* 2003;49:53–58.
- 39 Heinik J, Solomesh I, Berkman P: Correlation between the CAMCOG, the MMSE, and three Clock Drawing Tests in a specialized outpatient psychogeriatric service. *Arch Gerontol Geriatr* 2004;38:77–84.
- 40 Beinhoff U, Hilbert V, Bittner D, Grön G, Riepe MW: Screening for cognitive impairment: a triage for outpatient care. *Dement Geriatr Cogn Disord* 2005;20:278–285.
- 41 Kirby M, Denihan A, Bruce I, Coakley D, Lawlor BA: The Clock Drawing Test in primary care: sensitivity in dementia detection and specificity against normal and depressed elderly. *Int J Geriatr Psychiatry* 2001;16:935–940.
- 42 O'Rourke N, Tuokko H, Hayden S, Beattie BL: Early identification of dementia: predictive validity of the clock test. *Arch Clin Neuropsychol* 1997;12:257–267.
- 43 Lee H, Swanwick GRJ, Coen RF, Lawlor BA: Use of the clock drawing task in the diagnosis of mild and very mild Alzheimer's disease. *Int Psychogeriatr* 1996;8:469–476.
- 44 Casartelli M, Caputo L, Pesavento N, Lucchi T, Annoni G, Vergani C: The Clock Drawing Test as a screening instrument of cognitive function in the elderly. *Ric Psicol* 1995;19:87–108.
- 45 Manos PJ: Ten-point clock test sensitivity for Alzheimer's disease in patients with MMSE scores greater than 23. *Int J Geriatr Psychiatry* 1999;14:454–458.
- 46 Chan CC, Yung CY, Pan PC: Screening of dementia in Chinese elderly adults by the Clock Drawing Test and the time and change test. *Hong Kong Med J* 2005;11:13–19.
- 47 Cacho J, Garcia-Garcia R, Fernandez-Calvo B, Gamazo S, Rodriguez-Peres R, Almeida A, Contador I: Improvement pattern in the Clock Drawing Test in early Alzheimer's disease. *Eur Neurol* 2005;53:140–145.
- 48 Scanlan J, Borson S: The Mini-Cog: receiver operating characteristics with expert and naïve raters. *Int J Geriatr Psychiatry* 2001;16:216–222.
- 49 Heinik J, Solomesh I, Lin R, Raikher B, Goldray D, Merdler C, Kemelman P: Clock Drawing Test-modified and integrated approach (CDT-MIA): description and preliminary examination of its validity and reliability in dementia patients referred to a specialized psychogeriatric setting. *J Geriatr Psychiatry Neurol* 2004;17:73–80.
- 50 Brodaty H, Moore CM: The Clock Drawing Test for dementia of the Alzheimer's type: a comparison of three scoring methods in a memory disorders clinic. *Int J Geriatr Psychiatry* 1997;12:619–627.

- 51 Storey JE, Rowland JTJ, Basic D, Conforti DA: A comparison of five clock scoring methods using ROC (receiver operating characteristic) curve analysis. *Int J Geriatr Psychiatry* 2001;16:394–399.
- 52 Scanlan JM, Brush M, Quijano C, Borson S: Comparing clock tests for dementia screening: naïve judgements vs. formal systems – what is optimal? *Int J Geriatr Psychiatry* 2002;17:14–21.
- 53 Van der Burg M, Bouwen A, Stessens J, Ylief M, Fontaine O, de Lepeleire J, Buntinx F: Scoring clock tests for dementia screening: a comparison of two scoring methods. *Int J Geriatr Psychiatry* 2004;19:685–689.
- 54 Yamamoto S, Mogi N, Umegaki H, Suzuki Y, Ando F, Shimokata, Iguchi A: The Clock Drawing Test as a valid screening method for cognitive impairment. *Dement Geriatr Cogn Disord* 2004;18:172–179.
- 55 Leung JCW, Lui VWC, Lam LCW: Screening for early Alzheimer's disease in elderly Chinese patients using the Chinese Clock Drawing Test. *Hong Kong J Psychiatry* 2005;15:14–17.
- 56 Powlishta KK, Von Dras DD, Stanford A, Carr DB, Tsering C, Miller JP, Morris JC: The Clock Drawing Test is a poor screen for very mild dementia. *Neurology* 2002;59:898–903.
- 57 Seigerschmidt E, Mosch E, Siemen M, Förstl H, Bickel H: The Clock Drawing Test and questionable dementia: reliability and validity. *Int J Geriatr Psychiatry* 2002;17:1048–1054.
- 58 Connor DJ, Seward JD, Bauer JA, Golden KS, Salmon DP: Performance of three clock scoring systems across different ranges of dementia severity. *Alzheimer Dis Assoc Disord* 2005;19:119–127.
- 59 Chiu YC, Li CL, Lin KN, Chiu YF, Liu HC: Sensitivity and specificity of the Clock Drawing Test, incorporating Rouleau scoring system, as a screening instrument for questionable and mild dementia: scale development. *Int J Nurs Stud* 2008;45:75–84.
- 60 Moore MF, Wyke MA: Drawing disability in patients with senile dementia. *Psychol Med* 1992;14:97–105.