Simple scoring of the Clock-Drawing Test for dementia screening

Ejnar Alex Kørner[1], Lise Lauritzen[1], Flemming Mørkeberg Nilsson[2], Annette Lolk[3] & Peder Christensen[4]

ABSTRACT

INTRODUCTION: This study assessed five scoring methods of the Clock-Drawing test (CDT).

MATERIAL AND METHODS: A total of 72 out-patients and 29 healthy controls were assessed three times. At Visit 1, diagnostic procedure and assessments were performed with the Clinical Global Impressions (CGI) and Global Deterioration Scale (GDS), and the CDT and the Mini Mental State Examination (MMSE) were done blinded by a nurse. At Visit 2, CDT and MMSE were repeated, and at Visit 3 the CDT, CGI and the GDS were repeated. The CDTs were then rated by physicians and nurses using five different methods of scoring. Receiver-operating characteristics curve analyses were used to assess the CDT's suitability as a screening tool. Correlations between the five CDTs, other scales and predictive values were calculated. The extent to which three-word recall could improve the predictive values was analysed.

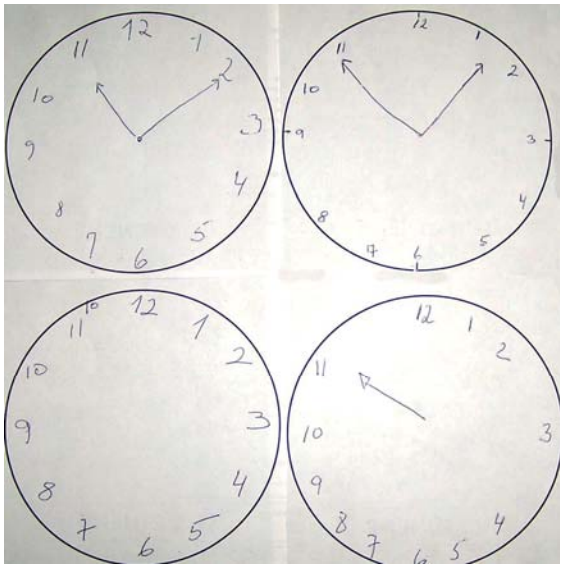
RESULTS: Correlations between the CDTs and the other scales were good. The predictive values were almost identical (positive values: 93-97%; negative values: 70-74%). Three-word recall improved the values. Rates of dementia in general practice and corresponding predictive values were estimated which resulted in markedly lower positive values around 60% for a rate of dementia of 20%, and 40% for a rate of dementia of 10%.

CONCLUSION: As predictive values were nearly identical, the shortest scoring manual (0 to 1) seems preferable.

FUNDING: The study was partly funded by Novartis Pharma A/S.

TRIAL REGISTRATION: Scientific Ethical Committee, 2003-2-17.

The Clock-Drawing Test (CDT) [1] is a cognitive test with a number of scoring variations, most of which are fairly easy and simple to perform and assess. Following the initial publication on the CDT, several clinicians working with dementia have published studies using different manuals for performing, scoring and interpreting the CDT [2]. No consensus on the most appropriate method for each clinical situation has ever been reached [1]. Nevertheless, the clinical use of the CDT has increased considerably over the past ten years. In Denmark, the CDT is part of a cognitive screening recommended for use in individuals applying for an extension of their driving permits beyond their seventieth birthday. The CDT together with questions on orientation and three-word recall form the cognitive test usually performed in the surgery of general practitioners (GPs) [3]. In the present study, we investigated the clinical validity of the CDT as a screening instrument for cognitive decline or dementia. This is the first study of its kind on a Danish material.



Four attempts at the Clock Drawing test. The upper left is correctly performed, the upper right and the lower left do not give the right time, while the lower right clock shows additional problems in placing the numbers correctly.

MATERIAL AND METHODS

Participants

The participants comprised patients referred to four psycho-geriatric out-patient services in Denmark; normal controls were recruited, mostly among the patients' caregivers and from local private organizations for the elderly. The patients had to fulfill the International Classification of Diseases (ICD)-10 criteria for dementia [4], while the controls were recruited only if these criteria were not met. All subjects were between 65 and 90 years of age. Candidates for participation were excluded if they suffered from aphasia, impaired hearing or sight severe enough to interfere with their ability to be assessed on the scales applied. For the same reason only Danish-speaking participants could be included.

ORIGINAL ARTICLE

1) Psychogeriatric Outpatient Clinic, Psychiatric Centre North Zealand, Mental Health Service, Capital Region
2) Psychogeriatric Outpatient Clinic 804, Psychiatric Centre Hvidovre
3) University of Southern Denmark, Institute of Clinical research, Research Unit of Psychiatry, Department P, Odense University Hospital
4) Psychogeriatric Department Haderslev, Mental Health Service, Southern Region of Denmark

Dan Med J 2012;59(1):A4365

The participants have been described in more detail elsewhere [5].

Scales and diagnosis

The following scales were applied

A: The Clock-Drawing Test (CDT) [1]: As part of a comprehensive set-up, the CDT was performed on a predrawn circle with a 10.6 cm diameter. Participants were asked to fill-in the numbers and set the time to ten past eleven. Participants were not allowed to look at another clock for guidance. In some cases, numbers were written outside the circle. As it became clear that this was a habit in certain trades and businesses, this particular variant was accepted as correct.

B. The Mini Mental State Examination (MMSE) has eleven items with a score from 0 to 30, a low score being indicative of cognitive deterioration. The MMSE version used in this study has been described elsewhere [6]. The items regarding orientation, three-word recall and the CDT are used as a cognitive screening test when elderly apply for an extension of their driver's permit beyond their seventieth birthday.

C. The Clinical Global Impression (CGI) [7] is a global scale used by trained clinicians to assess the severity of a particular condition. Scale scores range from 1 to 7; 1 = normal, not at all ill; 2 = borderline mentally ill; 3 = mildly ill; 4 = moderately ill; 5 = markedly ill; 6 = severely ill; 7 = among the most severely ill.

D. The Global Deterioration Scale (GDS) [8] assesses the degree of severity of dementia disorders. The scale has seven items: 1 = subjectively and objectively normal, independent; 2 = subjective complaints, objectively normal, independent; 3 = earliest signs of deficits, objective deficits, independent; 4 = clinically obvious deficits on clinical interview, may live independently; 5 = unable to survive without assistance, disorientation; 6 = will require assistance with basic activities of daily life, often in nursing home; 7 = incontinent, verbal activities lost, always in nursing home. The GDS is not a diagnostic scale; it is used once a diagnosis of dementia has been made.

Assessment programme

Data collection took place from March 2003 to August 2005. Twelve physicians and 16 nurses participated. Re-assessment of the CDTs made took place in 2007. All participants were assessed three times. At the first visit, a diagnosis using the ICD-10 criteria and assessments with the CGI and the GDS was made by one of

TABLE 1

The five manuals used for scoring the Clock-Drawing Test (CDT).

CDT 1 [9] score 0-5	CDT 2 [10] score 0-2	CDT 3 [11] score 0-1	CDT 4 [12] score 1-10	CDT 5 [13] score 1-6
5: Numbers and hands are correctly placed	2: Are numbers present and have they been placed correctly?	1: The clock drawing is correct	10-6: The drawing of a watch with circle and numbers that are usually intact	1: A clock has nearly been drawn
4: Mild visuo-spatial errors	1: Are the hands present and have they been placed correctly?	0: The clock drawing is not correct	10: Hands are correctly placed	2: Numbers are placed in the correct order, their placing is disregarded
3: Clear-cut errors in time given	0: Neither numbers nor hands are present or they have been placed incorrectly		9: Minor errors in placing of hands	3: Numbers are in correct order and correctly placed
2: Moderate visuo-spatial errors			8: More obvious errors in placing of hands	4: Hands are drawn, their placing is disregarded
1: Marked visuo-spatial errors			7: Placing of hands are completely incorrect	5: Hands show the approximately correct time, but are not placed quite correctly
0: Neither numbers nor hands are remotely correctly placed			6: Incorrect use of hands (e.g. circling the time, using digital numbers etc.)	6: Hands are placed correctly
			5-1: Drawing of watch, circle and numbers are not intact	
			5: Numbers are all placed at one end of the circle, numbers are reversed, etc., hands may still be present in some form	
			4: Placing of numbers is even more incorrect	
			3: Hands and numbers are no longer coherent. Hands may not be present	
			2: The drawing implies that parts of the instruction has been understood; however, only weak signs of a watch is recognizable	
			1: No attempt or only rudimentary attempts to draw a clock may be recognized	

the physicians. Following this and on the same day, participants were assessed by one of the nurses using the MMSE and the CDT as well as other scales [5]. On the second visit one week later, nurses repeated the MMSE and the CDT. The third visit took place six month after the first and the assessment programmes were identical. Blinding of the results was upheld between the physicians' and the nurses' test results throughout the study. Co-rating sessions were done to ensure the reliability of the test; nine co-ratings of the MMSE and CDT and ten of the CGI and GDS were held based on videotaped recordings of patients.

After completion of the primary study, copies of the original CDT results were made and distributed among the participating clinicians together with five different scoring instructions, **Table 1**, including: a modified version described by Shulman et al [9], the CDT as part of a short mental status test [10] and as part of the Mini-Cog (i.e. the CDT combined with the three-word recall test) [11], the 10-point version by Sunderland et al [12], and a version by Shua-Haim et al [13]. These assessments were performed independently and with no possibility of mutual interference.

Data analysis

Test-retest results and inter-rater reliability were analysed using intra-class-coefficients [14]. The CDT's value as a screening-tool was analysed using receiver-operating characteristics curve (ROC)-analyses with the ICD-10 diagnoses as the golden standards [15]. The most optimal cut-off value was decided for each CDT scoring, and each CDT was tested for correlation with other parameters such as the MMSE, the GDS and the CGI using the Spearmann correlation coefficients [16]. Positive and negative predictive values were calculated for the study population and estimated for the population in general practice. The base rates of dementia were 71% in all analyses; the prevalence of dementia in general practice is unknown. We chose to set it at 10% and 20%, the latter percentage being approximately the double of the estimated prevalence of dementia in the Danish population within the age range of the participants of this study, i.e. 11.5% [17]. It was also analysed how many of the false positive and false negative participants could subsequently be captured by the recall item of the MMSE as an add-on item using a cut off of one false answer. Furthermore, we analysed the correlation of the CDT and the item: "copying two overlapping pentagons" of the MMSE, as this item also assesses the visuo-spatial function.

Ethics

The study was partially funded by Novartis Pharma a/s. The study was performed in accordance with the Helsinki declaration and approved by the local scientific ethics committee. All participants received verbal and written information and written consent of participation was given. No other trial registration was needed.

Trial registration: Scientific Ethical Committee, 2003-2-17.

RESULTS

A total of 101 persons were included in the study, the age and gender distribution as well as the MMSE scores are illustrated in **Table 2**. In all, 29 were non-demented, 59 suffered from probable Alzheimer's disease, eight had probable vascular dementia and five had other forms of dementia disorders. Eighty-two (15 controls) were re-assessed at visit two, ninety at visit three; however, two were too cognitively impaired to fully participate in the testing which left 88 (27 controls) data sets for analysis. No controls were found to fulfil the dementia criteria at visit three. Statistically significant differences were found between the participating patients and controls regarding age and MMSE score. The intra-class-coefficients [10] of the MMSE, CGI and GDS ratings were all satisfactory (0.98, 0.88 and 0.69, respectively). The test-retest of the original CDT was satisfactory (0.74). The inter-observer reliability for all five sets of CDT scoring when used by the physician was almost perfect (0.98-0.99), while that of scoring set four was somewhat lower (0.89); however, this was still almost perfect when applied by the nurses.

The correlations between the five CDT sets and the CGI and GDS ranged from 0.69 to 0.79. The highest correlation was observed for the most specific scoring set (no. 4) and the lowest resulted from the least specific (no. 3). The correlation with the MMSE was somewhat weaker; however, it remained acceptable, ranging from 0.70 to 0.81, while the correlation with the copying of pentagons was weaker still, 0.63-0.69.

TABLE 2

The distribution of age and gender, Mini-Mental-State-Examination, Clinical Global Impressions and Geriatric Deterioration Score scores for the participants suffering from dementia and the healthy controls.

	Demented (n = 72)	Non-demented (n = 29)	p values[a]
Women, n (%)	45 (63)	17 (59)	NS
Age, years, mean ± SD	80.5 ± 8.3	75.3 ± 6.7	0.003
MMSE score, median (range)	17 (1-29)	29 (26-30)	0.0001
MMSE score, mean ± SD	15.6 ± 6.9	28.6 ± 1.5	0.0001
CGI, mean ± SD	4.4 ± 1.1	1.0 ± 0.0	0.001
GDS, mean ± SD	4.9 ± 1.8	1.2 ± 0.4	0.001

CGI = Clinical Global Impressions; GDS = Global Deterioration Scale;
MMSE = Mini Mental State Examination; NS = not significant; SD = standard deviation.
a) Wilcoxon signed rank; Mann-Whitney test.

TABLE 3

The most appropriate cut-off values of each Clock-Drawing Test. The positive and negative predictive values for each Clock-Drawing Test in the study and projected to artificial general practice populations based on rates of dementia arbitrarily set at 20% and 10%.

	CDT 1	CDT 2	CDT 3	CDT 4	CDT 5
Cut-off value	≤ 3	≤ 1	0	≤ 8	≤ 3
Sensitivity, %	86	87	85	87	86
Specificity, %	86	83	86	86	86
Study					
PPV, %	94	93	94	97	94
NPV, %	72	72	70	74	72
General practice 20%					
PPV, %	61	56	60	60	61
NPV, %	96	96	96	96	96
General practice 10%					
PPV, %	41	36	40	43	41
NPV, %	98	98	98	98	98

CDT = Clock-Drawing Test; NPV = negative predictive value; PPV = positive predictive value.

The results of the ROC analyses are given in **Table 3**. The optimal cut-off value for each scoring set is shown; it should be noted that scoring set no. 3 had only one possible cut-off value (i.e. the value 1) as it is dichotomous. Using these cut-off values, the predictive values were calculated for the study population. Predictive values were also calculated for prevalence rates closer to those likely to be found in general practice. Only small differences were found between the five scoring methods, in the study population as well as in the "general practice" population. The positive predictive values decrease and negative predictive values increase considerably as prevalence rates decrease. The number of false predictions in the study population varied between 13 (CDT 4) and 15 (CDT 3). When subsequently adding the recall item from the MMSE, the number of falsely predicted cases fell to five (CDT 2), four (CDT 3, 4 and 5) and three (CDT 1). In all CDTs, only one individual with dementia remained test-negative when the recall item was used.

DISCUSSION

The perfect scale for assessing dementia should be short and easy to administer. Furthermore, it should be applicable throughout the entire dementia disorder spectrum and it should reliably discriminate between demented and non-demented individuals.

Screening of dementia is often done by applying a number of scales and it has been customary that both the MMSE and the CDT were part of this set-up. The CDT to some extent assesses the frontal and temporo-parietal brain function by roughly screening the following cognitive abilities: understanding of verbal material, apraxia, visuo-spatial ability, executive functioning, and abstract thinking. The CDT may in this way be seen as supplementing the MMSE and it does not seem to emotionally affect the tested individuals [18]. The CDT seems to correlate with other cognitive tests and with the regional cerebral blood-flow in Alzheimer disease patients [19].

The CDT's adequacy as a screening tool has previously been studied [1, 20]. These studies did not find the CDT to be very reliable when screening for incipient and mild dementia, and they also criticised even earlier studies that reported satisfactory screening abilities for focusing on more advanced cases. When using the CDT with other cognitive tests, as has been done in this study, a general improvement in sensitivity has been shown; however, this was not achieved when the CDT was combined with the MMSE.

One limitation of the present study is the small number of controls and the fact that most of the participants with dementia were mildly to moderately ill persons referred for dementia assessment. The study therefore does not analyse the CDT's ability to discriminate between cognitively intact persons and persons with very mild impairment.

The advantages of this study are the fact that several centres participated in ensuring that individuals from rural areas, cities and from the Capital Region participated. The participating centres all diagnose and treat dementia disorders on a daily basis, which heightens the validity of the dementia diagnosis. On the other hand, it might be argued that a test intended for use in general practice ideally should be studied in this environment. In clinics such as those participating in the present study, dementia disorders are highly prevalent and the number of test-positives will be very high. The predictive values of a screening test studied under such circumstances will be overrated, and this must be taken into account when judging the CDT's clinical usability in everyday life in the GPs' practice. Albeit results should be interpreted with caution due to the small number of controls, this is corroborated by the calculations of the predictive values in "general practice" using dementia prevalence rates of 20% and 10%. Adding the recall item will reduce the false test-negative results, while false test-positive results remain largely unchanged. However, individuals with test-positive results should be referred for more thorough investigation at a memory or dementia clinic.

CONCLUSION

A single test that may decide whether an individual could safely continue to drive does not exist. Such evaluation depends on a number of factors, one of which is whether the individual suffers from a dementia disorder

requiring further examination and possibly treatment. In this study, we have tried to examine which scoring manual of the CDT is the best when screening for dementia. The differences in outcome between the individual scoring manuals are minor; all CDTs have a sensitivity and a specificity of around 86% and 87%. Positive predictive values ranged from 93% to 97%, while the negative predictive values ranged from 70% to 74 %. All enjoyed excellent inter-rater reliabilities. Even though the CDT 4, the most elaborate scoring manual, had slightly higher values, the CDT 3 seems to be the most recommendable owing to its simplicity. To increase the CDT's clinical usability, it is recommended to combine it with a three-word recall test.

CORRESPONDENCE: *Ejnar Alex Kørner,* Gerontopsykiatrisk Ambulatorium, Psykiatrisk Center Nordsjælland, 3400 Hillerød, Denmark.
E-mail: alex.koerner@regionh.dk

ACCEPTED: 3 November 2011

CONFLICTS OF INTEREST: Disclosure forms provided by the authors are available with the full text of this article at www.danmedj.dk.

ACKNOWLEDGEMENTS: The authors wish to thank the following for their contributions to this study: *August Wang, Carsten Schou, Kirsten Abelskov, Karen Vigsø, Christine Sweeney Hansen, Annette Brogaard,* Novartis Pharma A/S has sponsored meetings in relation to the study.

LITERATURE

1. Goodglass H, Kaplan E. The Assessment of aphasia and related disorders. Filadelphia: Lea and Fibiger, 1983.
2. Pinto E, Peters R. Literature review of the Clock Drawing Test as a tool for cognitive screening. Dement Geriatr Cogn Disorder 2009;27:201-13.
3. Hansen EA, Hansen BL. Cognitive functioning and driving ability in older drivers. Ugeskr Læger 2002;164:337-40.
4. WHO – ICD-10: Psykiske lidelser og adfærdsforstyrrelser. Klassifikation og diagnostiske kriterier. 1st ed. Copenhagen: Munksgaard Danmark, 1994
5. Kørner EA, Lauritzen L, Lolk A et al. The Neuropsychiatric Inventory-NPI. Validation of the Danish version. Nord J Psychiatry 2008;62:481-5.
6. Kørner EA, Lauritzen L, Nilsson FM et al. Mini mental state examination. Validation of new Danish version. Ugeskr Læger 2008;170:745-9.
7. Guy W. ECDEU assessment manual of psychopharmacology – revised (DHEW Publ no ADM 76 – 338) Rockville MD: Department of Health, Education and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, NIHM Psychopharmacology Research Branch, Division of Extramural Research Programs, 1976:218-22.
8. Reisberg B, Ferris SH, de Leon MJ et al. The Global Deterioration Scale (GDS) for assessment of primary degenerative dementia evaluation. Am J Psychiatry 1982;139:1136-9.
9. Shulmann KI, Gold DP, Cohen CA et al. Clock-Drawing and dementia in the community: a longitudinal study. Int J Geriatr Psychiatry 1993;8:487-96.
10. Kokmen E, Naessens JM, Offord KP. A short test of mental status: description and preliminary results. Mayo Clin Proc 1987;62: 281-8.
11. Borson S, Scanlan J, Brush M et al. The mini-cog: a cognitive "vial signs" measure for dementia screening in multi-lingual elderly. Int J Geriatr Psychiatry 2000;15:1017-21.
12. Sunderland T, Hill JL, Mellow AM et al. Clock drawing in Alzheimer's disease: a novel measure of dementia survey. JAGS 1989;37:725-9.
13. Shua-Haim J, Koppuzha G, Gross J. A simple scoring system for clock drawing for patients with Alzheimer's disease. J Am Ger Soc 1996;44:335.
14. Bartko JJ, Carpenter WT, Jr. On the methods and theory of reliability. J Nerv Ment Dis 1976;163:307-17.
15. Beck RJ, Schultz EK. The use of receiver operating characteristic (ROC) curves in test performance evaluation. Arch Pathol Lab Med 1986;110:13-20.
16. Altman DG. Spearmann's rank correlation coefficient r. In: Altman D.G.: Practical statistics for medical research. London: Chapman and Hall, 1991:295-6.
17. Andersen K, Lolk A, Nielsen H et al. Prevalence of very mild to severe dementia in Denmark. Acta Neurol Scand 1997;96:82-7.
18. Burns A, Lawlor B, Craig S. Clock Drawing Test. In: Burns A, Lawlor B, Craig S, eds. Assessment scales in old age psychiatry. London: Martin Dunitz Ltd, 1999:44.
19. Ueda H, Kitabayashi Y, Narumoto J et al. Relationship between clock drawing test performance and regional blood flow in Alzheimer's disease: a single proton emission computed tomography study. Psychiatry Clin Neurosci 2002;56:25-9.
20. Connor JC, Seward JD, Bauer JA et al. Performance of three clock scoring systems across different ranges of dementia severity. Alzheimer Dis Assoc Disord 2005;19:119-27.