The Clock Drawing Test: performance among elderly with low educational level

Abstract

Objective: To compare the accuracy of four scales, and to determine the inter-rater reliability and the influence of schooling on Clock Drawing Test performance in screening for dementia in elderly with low educational level. Method: Design: Criterion validation, concurrent type. Setting: Outpatient unit of Universidade do Estado do Rio de Janeiro Hospital, Brazil. Participants: A total of 306 individuals (≥ 65 years) were assessed by a geriatrician and a neuropsychologist using functional and cognitive instruments. The diagnosis of dementia was established by consensus in a meeting between these professionals. Four different researchers blinded as to the cognitive status of the patients scored the Clock Drawing Test using Shulman’s, Sunderland’s, Manos’, and Wolf-Klein’s methods; and a fifth researcher assessed Manos’ method for an inter-rater reliability study. Results: The data of 211 individuals (≤ 4 years of schooling) were analyzed. The sensitivity and specificity of the four methods were similar (about 65%). The inter-rater reliability of Manos’ method was excellent (ICC = 0.944). The areas under the ROC curves were small (0.657 to 0.681), and the differences between them were not statistically significant. Conclusions: The Clock Drawing Test is a reliable but not valid test to screen for dementia in older outpatients with 4 or less years of formal schooling.

Descriptors: Neuropsychology; Tests; Mass screening; Aging; Ambulatory care

Resumo

Objetivo: Comparar a acurácia de quatro escalas de avaliação e determinar a confiabilidade inter-aferidor e a influência da escolaridade, no desempenho do Teste do Desenho do Relógio para detecção de demência em idosos com baixo nível educacional. Método: Delineamento de estudo: Validação de critério, tipo concorrente. Local de estudo: Ambulatório do Hospital da Universidade do Estado do Rio de Janeiro, Brasil. Participantes: 306 indivíduos (≥ 65 anos) submetidos à avaliação funcional e cognitiva por um geriatra e um neuropsicólogo; uma reunião de consenso entre estes profissionais avaliou a presença de demência. Quatro pesquisadores, que desconheciam o status cognitivo dos indivíduos, pontuaram o Teste do Desenho do Relógio – métodos de Shulman, Sunderland, Manos e Wolf-Klein, e um quinto pesquisador repetiu o método de Manos para o estudo de confiabilidade inter-aferidor. Resultados: 211 indivíduos (≤ 4 anos de escolaridade) tiveram seus dados analisados. A sensibilidade e a especificidade dos quatro métodos foram similares e variaram em torno de 65%. A confiabilidade inter-aferidor foi excelente (ICC = 0,944). As áreas sob as curvas ROC foram pequenas (0,657 a 0,684) e as diferenças entre elas não foram estatisticamente significativas. Conclusões: O Teste do Desenho do Relógio é um instrumento confiável, mas não é válido para o rastreio de demência em pacientes idosos com quatro anos ou menos de escolaridade.

Descritores: Neuropsicologia; Testes; Programas de rastreamento; Envelhecimento; Assistência ambulatorial

Correspondence
Prof. Dr. Roberto Alves Lourenço
Policlínica Piquet Carneiro, Cuidado Integral à Pessoa Idosa
Universidade do Estado do Rio de Janeiro
Av. Marechal Rondon, 381 – 2º andar
20950-000 Rio de Janeiro, RJ, Brazil
E-mail: roberto.lourencoc@globo.com

1 Internal Medicine Department, Medical Sciences School, Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro (RJ), Brazil
2 Open University for Studies on the Elderly, Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro (RJ), Brazil

Submitted: December 13, 2007
Accepted: August 11, 2008
Introduction

In Brazil, the growing prevalence of chronic degenerative diseases is a consequence of one of the fastest demographic and epidemiological transitions in the world. In less than 50 years, the Brazilian population profile has changed dramatically, and it has been estimated that the proportion of elderly people will continue to rise very rapidly in the next decades.1 Since dementia is a common problem in the geriatric population, the already large medical and social costs of its treatment will predictably increase enormously. Brazilian epidemiological studies2 have found prevalence of dementia in community samples to be 7.1 % in a small urban area in the State of São Paulo,3 and 9.8% in the city of Rio de Janeiro, where the present study was conducted.4

Recent therapeutic advances in the area of dementia treatment, along with evidence suggesting that earlier treatment leads to greater benefits for the patient and his family,5 and to lower costs,6 have underscored the importance of detecting dementia in its initial stages.7

This early identification is inevitably largely under the generalist’s responsibility. However, studies have shown that, at the primary care level, it is underdiagnosed.8 This is possibly due to the relative reluctance of these professionals to apply cognitive screening tests.9

To overcome this barrier it is necessary to develop tests that are easily applicable, require short time (maximum of 5 minutes), are acceptable to the patients and health care professionals, and have good sensitivity and specificity for the diagnosis of dementia, even in populations with heterogeneous educational levels.

The Clock Drawing Test (CDT)10-14 satisfies most of these conditions, and is considered relatively resistant to educational and ethnic diversity, although the available data do not support this observation conclusively.15,16 Many studies have shown that the CDT is useful as a screening test for cognitive impairment, especially in populations with higher educational levels (8 years or more).17,18 It also has various application methods, each with its own scoring system, but no consensus as to which is the best has been obtained.

The validity of the CDT as a screening instrument for cognitive disturbance in the elderly has been the subject of numerous studies in the international literature.15-21

In Brazil, however, only Okamoto22 and Fuzikawa et al.23,24 have produced the few studies published about this instrument. Okamoto22 studied spontaneous clock drawings, ability to tell time, placement of the hour hand, and copies of clock drawings in 50 patients with Alzheimer’s dementia (DAT), nine with Lewy body dementia (DBL), six with mild cognitive impairment (MCI), and 43 from a control group. For this author, the CDT proved to be an adequate instrument to differentiate controls from those with DAT and DBL, although there were no differences between the DAT and DBL groups; he also concluded that the association of the Mini Mental State Exam (MMSE)25 with CDT subitems could improve the MMSE accuracy as a screening test.

Fuzikawa et al.23 assessed a random sample of 202 elderly individuals with very low formal educational level. The objective of this population-based cohort study was to determine intra- and inter-rater reliabilities of the CDT scored by the Shulman12 method; the elderly were scored on two occasions by the same rater and by two independent raters. In the authors’ opinion the method appeared to have good to excellent reliability in this elderly population.

Fuzikawa et al.24 evaluated a sample of 1,118 elderly subjects from a population-based cohort study to determine the correlation and agreement between the MMSE25 and CDT. Correlation between tests was moderate. Subjects who performed well on the CDT could be expected to obtain high MMSE scores. They concluded that, although one test does not substitute the other, the CDT may be more practical in developing countries where resources are limited and low education is common.

Despite the importance of these papers in exploring the first data about CDT performance in Brazilian people, the high schooling of the sample in the first study23 and the absence of a confirmatory standard in the last two23,24 limited our ability to conclude that the CDT is a useful tool to screen elderly individuals with low educational level for dementia.

The objectives of the present study were: 1) to compare the accuracy – sensitivity, specificity, and area under the curve – of four CDT scoring systems (Manos,11 Shulman,12 Wolf-Klein,13 Sunderland,14) 2) to determine the inter-rater reliability of one of these methods (Manos;11) and 3) to determine the influence of schooling on the CDT accuracy as a screening test for dementia in a population of older people treated in a primary care outpatient clinic, where low educational levels predominate.

Method

1. Sample selection

Between April 8th and July 15th, 2002, a convenience sample of 306 individuals was selected among older people who were seeking general medical treatment at the Internal Medicine Clinic of Policlinica Piquet Carneiro, an outpatient unit of Universidade do Estado do Rio de Janeiro Hospital. Individuals older than 65 years who attended our screening center seeking a general physician’s office were invited to participate in a study to validate instruments to screen for dementia.26 The number of participants recruited daily depended on how many accepted the invitation, and was limited by the absorption capacity of the research team at the time.

Inclusion criteria were age over 65 years and preserved hearing and comprehension, at least enough to participate in the study and sign an informed consent form. Exclusion criteria were report – personal or through an informant – of a serious uncorrected visual or auditory deficiency; being in an advanced stage of cognitive disturbance, or having any mental illness that could compromise understanding and performance on test procedures; native language other than Portuguese; difficulty in moving the hands due to rheumatic or neurological diseases.

After signing the informed consent form, the subjects were referred to a comprehensive geriatric evaluation, implemented by a multiprofessional team – a geriatrist, a registered nurse practitioner, a social worker, and a neuropsychologist. At the end of the evaluation, a meeting between the geriatrist and the neuropsychologist classified the patients into one of two dementia syndrome-based groups: demented and not demented. Therefore, the confirmatory standard for the diagnosis of dementia was the consensual opinion of both professionals, which took into account both the clinical impression and the neuropsychological evaluation, and was based on the DSM-IV27 diagnostic criteria for dementia syndrome. The patients were not further classified as to the cause of their dementia.

The Universidade do Estado do Rio de Janeiro Hospital Ethics Committee approved the research protocol, including the informed consent form. The study was supported entirely by the Universidade do Estado do Rio de Janeiro.
2. Clock Drawing Test and study tests and procedures

In addition to the clinical assessment algorithms pertaining to each of the specialties involved, the individuals were submitted to a functional evaluation, which included the Activities of Daily Living Scale\(^{28}\) and the Instrumental Activities of Daily Living Scale;\(^{29}\) to the Geriatric Depression Scale;\(^{30}\) and to the MMSE.\(^{24}\)

The neuropsychometric tests included the cognitive part of the Cambridge Examination for Mental Disorders of the Elderly - Revised (CAMDEX-R),\(^{31}\) the CAMCOG-R. This is a concise neuropsychological test battery for the assessment of cognitive impairment in elderly people, and examines a broad range of cognitive functions, such as memory, language, attention, perception, and executive functions. A clock drawing task is part of this instrument and uses the following instruction: “Draw the face of a large clock, place all the numbers inside and place the pointers indicating 11:10 (eleven hours and ten minutes)”. The clock drawing of each patient – copied and identified with its respective register number – was analyzed retrospectively by five researchers who had no access to the patient’s file and were unaware of the cognitive condition of the subjects; each researcher applied one of the following four methods: Manos,\(^{11}\) Shulman et al.,\(^{12}\) Wolf-Klein et al.,\(^{13}\) and Sunderland et al.\(^{14}\) proposals. The fifth researcher applied the Manos method,\(^{11}\) and thus this test was assessed by two researchers with the aim of evaluating its inter-rater reliability. The researchers standardized these different methods by three procedures: 1) reading of the original texts with the description of the items and subsequent studies that used them; 2) intense discussion about the operational aspects to establish standardization in the instrument’s application; 3) a pilot project with 35 clock drawings which confirmed that the standardization of these four methods among the members of the research team was excellent. The intraclass correlation coefficient (ICC) results were: 0.96 (95% CI 0.88-0.98); 0.95 (95% CI 0.90-0.98); 0.89 (95% CI 0.79-0.95); 0.97 (95% CI 0.95-0.99) for the Manos,\(^{11}\) Shulman et al.,\(^{12}\) Wolf-Klein et al.,\(^{13}\) and Sunderland et al.\(^{14}\) methods, respectively.

As to the choice of scoring systems, the most frequently cited in the literature were selected. The method proposed by Wolf-Klein et al.\(^{13}\) uses a table of images as a guide to classify and score the different error patterns, and the scores vary from 1 to 10 (best) points; the Manos\(^{11}\) proposal uses a transparent pre-drawn mask that is placed over the clock made by the patient, generating a very objective and simple scoring system, with results ranging from 0 to 10 (best). The methods by Sunderland et al.\(^{14}\) and Shulman et al.\(^{12}\) use descriptive tables, with the scores expressed as 1 to 10 (best) and 1 to 6 (worst), respectively. The CDT cut-off points originally used by Wolf-Klein et al.,\(^{13}\) Manos;\(^{11}\) Sunderland et al.\(^{14}\) and Shulman et al.\(^{12}\) to screen elderly for dementia are 6/7, 7/8, 5/6, and 3/2, respectively.

3. Statistical analysis

ROC curves, areas under the curves and their confidence intervals were plotted, and the best trade-offs between sensitivities and specificities were calculated for each of the four methods. The areas under the ROC curves were compared using the z-statistics. The Kolmogorov-Smirnov test for normality was used to test variables for normal distribution. Comparisons between categorical variables were performed using Pearson’s chi-square test, and the Mann-Whitney U test or, when applicable, the Kruskal-Wallis test was used to compare means of not normally distributed variables. The inter-rater reliability was measured as an ordinal variable by the intraclass correlation coefficient. The data were entered and analyzed using SPSS for Windows, version 11. The MedCalc statistical program was used to compare the areas under the ROC curves. A significance level of 5% was adopted for all tests.

Results

A total of 306 subjects were recruited; 293 completed all the study procedures; 211 had 4 or less years of schooling and had their data analyzed. Of these, 153 (72.5%) were female, 59.7% declared they did not live with a partner, 82.5% were not working, 64.9% were retired, and 37% declared they had never attended school. In addition, 71.1% were under 75 years and 12.3% were more than 80 years of age (Table 1). According to the DSM-IV,\(^{27}\) 32.2% fulfilled criteria for dementia syndrome. Age varied between

| Table 1 - Socioeconomic characteristics stratified by the diagnosis of dementia – DSM-IV\(^{27}\) (n = 211) |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| Items            | Dementia (n%)   | Total (n%)      | p-value         |
| Gender           | Male            | 16/27.6         | 58/27.5         | 0.374           |
|                 | Female          | 52/32.4         | 153/72.5        |                 |
| Age (yrs)        | 65-69           | 15/24.6         | 61/28.9         |                 |
|                 | 70-74           | 29/42.6         | 89/42.2         |                 |
|                 | 75-79           | 11/18.6         | 35/16.6         |                 |
|                 | ≥ 80            | 13/20.0         | 26/12.3         | 0.144           |
| Schooling (yrs)  | 0               | 34/43.6         | 78/37.0         |                 |
|                 | 1               | 8/12.6          | 25/11.8         |                 |
|                 | 2               | 11/16.4         | 18/8.1          |                 |
|                 | 3               | 13/19.1         | 41/19.1         |                 |
|                 | 4               | 7/14.0          | 50/23.7         | 0.001           |
| Marital Status   | Single          | 6/30.0          | 20/9.5          |                 |
|                 | Separated       | 8/32.0          | 25/11.8         |                 |
|                 | Widow           | 53/65.4         | 81/38.4         |                 |
|                 | Married         | 55/69.6         | 79/37.4         | 0.946           |
| Work             | Yes             | 3/10.0          | 30/14.2         |                 |
|                 | No              | 11/53.8         | 174/82.5        | 0.004           |
| Retirement       | Yes             | 95/69.0         | 137/64.9        |                 |
|                 | No              | 44/63.8         | 69/32.7         | 0.420           |

*DSM-IV = Diagnostic and Statistical Manual of Mental Disorders, 4th Edition*
65 and 93 years, and its mean (± SD), median and mode were 72.8 (± 5.4), 72, and 73, respectively; the mean ages were 74.0 (± 5.8) and 72.0 (± 5.0) for demented and non-demented groups, respectively. The mean, median and mode number of years of schooling were 1.8 (± 1.7), 2, and 0; the means and median were 1.28 (± 1.49) and 0.5, and 2.06 (± 1.66) and 2 for demented and non-demented groups, respectively. As can be seen in Table 1, there was a statistically significant difference between demented and non-demented groups in the variables “Work” and “Schooling.”

Table 2 presents the CDT sensitivities and specificities when assessed by the four different methods, according to the original cut-off points proposed by the authors and the best cut-off points – trade-offs between sensitivity and specificity – seen in the present study. The sensitivities and specificities varied around 65%.

There was also no clear superiority of one method of CDT evaluation over the others. Table 3, Table 4, and Figure 1 show the areas under the curves, the comparison between them, and the profiles of ROC curves, with their respective 95% confidence intervals, standard errors, and significance levels. The comparison between these areas shows no significant differences (Table 4). Although the performance of the four methods has improved progressively with each additional year of schooling (p = 0.000), excluding the group of illiterates does not improve the overall accuracy of the CDT, irrespective of which scoring method is used.

The inter-rater reliability of Manos’ method was considered excellent: ICC = 0.944.

Discussion

The development of dementia screening tests that are easily applicable by generalists and have adequate accuracy, even in populations with low educational level, should be included in the objectives of programs for investigation of cognitive disturbances in older people.

Application of the MMSE,25 probably the most widely used test in primary care, lasts at least 10 minutes, which is relatively long in this context. In addition, it requires specific training by the professional who applies it and is strongly influenced by the educational level of the subjects.31-33 The CDT, on the other hand, is simple, fast (1 to 2 minutes), well accepted by the patients and health care professionals, and does not require prolonged previous training to be applied.10-15 In this respect, Scanlan et al. have shown that naïve raters, i.e. with no training at all, are able to classify CDT with the same accuracy as that obtained by trained raters.34 However, few studies have specifically addressed its psychometric characteristics in populations of low educational level.35,36 The results of our present study suggest that none of the methods tested have good accuracy in tracking dementia in a primary care setting among older people with 4 or less years of formal education, given that the calculated sensitivities and specificities were low.

Other authors have also documented the adverse effect of low schooling levels on the CDT performance.35,36 Storey et al. tested the CDT in a sample with a lower mean educational level and also found significant differences in performance for the subgroup with less than 6 years of schooling.35 Ostroski-Solis et al. have recently demonstrated that education exerts a very strong effect on neuropsychological tests in general, and that the educational variable is more significant in groups with the lowest levels of schooling, especially in those with less than 5 years of formal education.37

The cognitive abilities necessary to draw a clock are thought to include visuospatial and constructional skills, abstract reasoning, and executive functions, although many authors have used clock drawing as a test of visuocognitive abilities, executive and praxic functions, while others have preferred clock reading or clock setting as a means to evaluate symbolic representation.38

There is experimental evidence that even these skills are, in fact, dependent upon educational level. Ardila et al. compared the performance of illiterates and educated professionals in a wide variety of neuropsychological tests and found that almost all of the abilities tested were strongly influenced by education.39 Another study revealed differences in clocks drawn by non-demented adults of high education (> 8 years) and low education (≤ 8 years), suggesting that differences in schooling can cause differences in the drawings, regardless of the cognitive status.40

It is not necessary to be able to draw a clock face in order to understand and manipulate concepts related to the passage of time. This can explain, in part, the large number of people who, although not demented, are unable to produce “normal” clock drawings. However, these same people may be capable of telling time on a clock, or even of placing the pointers on a clock’s partially drawn face.41 This can also explain the low accuracy of the CDT in individuals with very low formal educational attainment, such as those included in the present study.

The inter-rater reliability results were comparable with the results of Fuzikawa et al.23 In fact, it seems that the level of education does not affect the instrument stability, irrespective of the method of analysis employed to classify the clock drawings.

Similarly, the results of Fuzikawa et al.42 suggest that there is a good correlation between the CDT and other cognitive tests (i.e., MMSE), even in illiterate elderly with low educational level. This is likely due to some degree of parallelism between these tools, in that they are probably measuring aspects of related cognitive constructs, but it does not necessarily follow that these tests have good accuracy.

Table 3 - Area under curve, standard error, and 95% confidence interval of the Manos,11 Shulman et al.,12 Wolf-Klein et al.,13 and Sunderland et al.14 methods of Clock Drawing Test evaluation

<table>
<thead>
<tr>
<th>Method</th>
<th>AUC*</th>
<th>SE*</th>
<th>95% CI*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manos11</td>
<td>0.668</td>
<td>0.038</td>
<td>0.600 to 0.731</td>
</tr>
<tr>
<td>Shulman et al.12</td>
<td>0.684</td>
<td>0.0408</td>
<td>0.617 to 0.746</td>
</tr>
<tr>
<td>Wolf-Klein et al.13</td>
<td>0.684</td>
<td>0.0372</td>
<td>0.617 to 0.746</td>
</tr>
<tr>
<td>Sunderland et al.14</td>
<td>0.657</td>
<td>0.0385</td>
<td>0.588 to 0.720</td>
</tr>
</tbody>
</table>

* AUC = area under the curve; SE = standard error; CI = confidence interval

Table 2 - Sensitivity and specificity of the Manos,11 Shulman et al.,12 Wolf-Klein et al.,13 and Sunderland et al.14 methods of Clock Drawing Test evaluation according to original cut-off points proposed by the authors and the best cut-off points seen in the present study

<table>
<thead>
<tr>
<th>Method</th>
<th>Sensitivity cut-off</th>
<th>Specificity cut-off</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Original</td>
<td>Best</td>
</tr>
<tr>
<td>Manos11</td>
<td>63%</td>
<td>88%</td>
</tr>
<tr>
<td>Shulman et al.12</td>
<td>65%</td>
<td>84%</td>
</tr>
<tr>
<td>Wolf-Klein et al.13</td>
<td>62%</td>
<td>62%</td>
</tr>
<tr>
<td>Sunderland et al.14</td>
<td>59%</td>
<td>59%</td>
</tr>
</tbody>
</table>

* Best cut-off: Manos: 4/5; Shulman et al.: 4/3; Wolf-Klein: 5/6; Sunderland: 5/6
* Original cut-off: Manos: 7/8; Shulman et al.: 3/2; Wolf-Klein: 6/7; Sunderland: 5/6

Rev Bras Psiquiatr.
Although good reliability and correlation are necessary conditions for validity, these characteristics are not sufficient to reach a conclusion as to the instrument’s capacity to detect suspicion of dementia. This could be a good reason for justifying the development of study designs such as that used in this study, which aim at directly measuring the CTD accuracy in elderly individuals with very low educational.

Higher age and lower educational level have been found to be risk factors associated with dementia syndrome. The results show a non-significant (p-value = 0.144), but visible, trend to higher frequencies of dementia with aging, and a significant (p = 0.001) tendency for its manifestation in the elderly with the worst levels of schooling (Table 1).

This study has some limitations. The first refers to the fact that, for most of the patients, there is no complete information from third parties (caregivers, relatives) on previous functional and cognitive status, which may have caused some misclassification errors. The second is that the clock drawings used were part of the CAMCOG-R, the neuropsychological test included in our diagnostic criteria algorithm. Although the CDT was only a fragment of the whole test, this information was included as part of that used to establish the diagnosis of dementia, and it could have produced some information bias.

Other limitations to be considered are associated with the method of sample selection. To adequately address selection bias, a randomized sample would have been better than the convenience sample used in the present study. In the latter, the true dementia prevalence in the study population may be misrepresented, and this is a factor which could interfere with the psychometric characteristics of the instrument.

On the other hand, our sample was composed of individuals who were seeking primary health care, having signs and symptoms not necessarily related to cognitive disorders. This probably reduced the problem of selection bias, frequently seen in studies that select their samples by the diagnosis condition of the subjects, with the risk of inflating test accuracy.

A second positive aspect was that the sample had a large proportion of older people with low educational level, which permitted direct calculations of their performance on the instrument being tested. Finally, although the present study was not intended to produce knowledge about the CDT scale characteristics in Brazilian populations with higher educational levels (> 5 years), such a task must be implemented since there are no definitive studies on this specific group.

In conclusion, the CDT is strongly influenced by educational level, and although there is some evidence from the international and Brazilian literature that it can be used to screen older people with higher educational levels, it seems that it is not adequate for dementia screening in individuals with less than 5 years of formal education; in addition, the four scoring methods tested had similar accuracy to screen older outpatients for dementia.

**Acknowledgments**

Conflict of interest declaration: The authors state that they have no financial disclosures or conflict of interests to report on any aspect of the items reported here.

Authors contributions: Dr Roberto A. Lourenço was responsible for the overall design of the study, assessment of clock drawings, analysis and discussion of results, and paper writing; Dr Sergio T. Ribeiro- Filho collaborated in the assessment of clock drawings, analysis and discussion of results, and paper writing; Irene F. H. Moreira, Emylucy M. P. Paradela, and Aline S. Miranda participated in the assessment of clock drawings, discussion of results, and paper writing.
References


