Published in final edited form as:
J Neuropsychiatry Clin Neurosci. 2010 ; 22(1): 85–92. doi:10.1176/appi.neuropsych.22.1.85.

Clock Drawing Test Ratings by Dementia Specialists: Interrater Reliability and Diagnostic Accuracy

Anil K. Nair, M.D., **Brandon E. Gavett, Ph.D.**, **Moniek Damman**, **Welmoed Dekker**, **Robert C. Green, M.D., M.P.H.**, **Alan Mandel, M.D.**, **Sanford Auerbach, M.D.**, **Eric Steinberg, M.S.N., A.P.R.N., B.C.**, **Emily J. Hubbard, M.P.H.**, **Angela Jefferson, Ph.D.**, and **Robert A. Stern, Ph.D.**
Alzheimer's Disease Center, Department of Neurology, Boston University School of Medicine

Abstract

The authors aim to study subjective ratings of clock drawing test by clinicians and determine interrater reliability and diagnostic accuracy. The clock drawing test has been advocated over the Mini-Mental State Examination as an office screening test for dementia, but use of the clock drawing test by neurologists and dementia specialist clinicians has not been validated. The authors conducted a study of clock drawing test scoring by dementia specialists. The authors randomly assigned 25 clocks from each of six predetermined groups based on consensus diagnosis (cognitive comparison subjects, subjects with a memory complaint but with normal neuropsychological testing, subjects with probable and possible mild cognitive impairment, and subjects with possible and probable Alzheimer's disease) to dementia specialists for blinded scoring using a binary yes/no impairment system and a 0–10 scale as subjectively determined by each individual clinician rater. The authors collapsed the six groups into three (comparison subjects, mild cognitive impairment patients, and Alzheimer's disease patients) and analyzed interrater reliability, sensitivity, and specificity for consensus diagnosis of mild cognitive impairment, and Alzheimer's disease. The authors found excellent interrater reliability, sensitivity, and specificity for predicting consensus diagnosis. The 0–10 clock drawing test rating scale was more predictive of consensus diagnosis than the binary impairment scale. Based on the five clinicians' average dichotomous rating, the clinicians differentiated comparison and Alzheimer's disease participants with a sensitivity of 0.75 and a specificity of 0.81. For three of the four comparisons, a cutoff score of two or greater resulted in the maximization of sensitivity and specificity for differentiating diagnostic groups. A cutoff score of four or greater maximized sensitivity (0.54) and specificity (0.74) for differentiating Alzheimer's disease from mild cognitive impairment. Based on rating systems, clock drawing test scoring by dementia clinicians had excellent interrater reliability and sensitivity for differentiating the mild Alzheimer's disease subjects from comparison subjects. When utilizing a binary rating scale for the clock drawing test in the absence of clinical information, dementia specialist clinicians at the Boston Medical Center were moderately sensitive and highly specific in separating mild cognitive impairment from healthy comparison subjects. These dementia clinicians were also highly sensitive and less specific in differentiating mild cognitive impairment from Alzheimer's disease.

Alzheimer disease is a growing public health problem[1,2] and its prevalence is increasing rapidly with the aging of the baby boomer generation.[3] Early diagnosis and treatment can reduce the burden this increase poses to the health care system and society.[4] However, Alzheimer's

Address correspondence to Anil K. Nair, M.D., Assistant Professor of Neurology, 715 Albany St., B7800, Boston, MA 02118; an@bu.edu.

A portion of this work was presented in poster format at the American Academy of Neurology Annual Meeting, April 14–16 2008, in Chicago. None of the authors have any conflicts of interest to disclose. There are no commercial associations that might pose or create a conflict of interest with the information presented in the submitted manuscript.

disease is often under recognized in community clinical practice settings[5–7] because the diagnosis can be difficult[8] and may require specialized training. Without fast and reliable screening instruments, it may be difficult for primary care physicians to identify patients who should be referred for a more comprehensive dementia workup.

The clock drawing test is widely used as a screening test for dementia. Neurologists have used clock-drawing and time telling tests extensively.[9] Several factors contribute to the test's popularity, including administration and scoring ease and evaluation of multiple cognitive domains,10,11 such as executive functioning.11–13 Compared to the Mini-Mental State Examination (MMSE), the clock drawing test is thought to have less educational bias[14] and is better able to detect cognitive decline due to Alzheimer's disease and other dementias.15 The clock drawing test has also been advocated over the MMSE as an office screening test for dementia in community clinics4 and in acute hospital settings.16 Furthermore, the clock drawing test is suitable for non-English speaking populations.[14]

There are two general clock drawing test scoring approaches, including qualitative and quantitative, and varied scoring systems that emphasize different facets of the clock drawing process. Early quantitative scoring systems were validated to distinguish between subjects with moderate or severe Alzheimer's disease and cognitive healthy comparison subjects and later adapted for use in mild cognitive impairment and mild Alzheimer's disease.[17–23] Previous studies of objective clock drawing test rating systems identified Alzheimer's disease with overall diagnostic accuracy ranging from 59% to 85%.24 However, such diagnostic accuracy has not been found in mild cognitive impairment cohorts with sensitivities ranging from 17% to 92%.24 In a retrospective study comparing several clock drawing test scoring systems,24 the scoring system by Mendez[19] has been found to be the most accurate in distinguishing demented from nondemented individuals, followed closely by the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) system.[25]

Though diagnostically useful, quantitative clock drawing test rating schemes[18,19,20,22,23] are rarely used in clinical settings, as they take more time and require trained clinical personnel to score. Moreover, when using the clock drawing test to identify dementia, qualitative ratings of naive judges may be equal to or more accurate than many quantitative scoring systems.[24] Though it is widely assumed that dementia specialists are more reliable and valid than naive raters, there are no known studies that have evaluated the psychometric properties of clock drawing test ratings made by trained clinicians who use the clock drawing test as part of their regular clinical practice. Our present study was performed to determine the interrater reliability, sensitivity, and specificity of qualitative clock drawing test ratings made by clinicians specializing in the assessment of patients with dementia. Two qualitative rating approaches were utilized: a dichotomous rating of impaired versus nonimpaired and a 0–10 ordinal rating scale. A multidisciplinary consensus conference was the gold standard for dementia diagnosis in the current study.

METHODS

Participants

Archival data were extracted from the Boston University Alzheimer's Disease Core Center registry, which is an institutional review board approved, National Institute on Aging-funded, Alzheimer's disease registry,26–28 that longitudinally follows older adults with and without memory problems. Participants performed the clock drawing test as part of an annual neurological and neuropsychological exam. All participants were at least 55 years old, English-speaking community dwellers, with no history of major psychiatric or neurological illness or head injury involving loss of consciousness, and with adequate auditory and visual acuity to complete the examination. After data query, there were 506 eligible participants in the Boston

University Alzheimer's Disease Core Center patient/comparison registry who had been diagnosed by a multidisciplinary consensus team (including at least two board-certified neurologists and two neuropsychologists) based on a clinical interview with the participant and an informant, medical history review, and neurological and neuropsychological examination results. Of the 506 participants, 168 were diagnosed as cognitively normal comparison subjects, 39 as cognitively normal comparison subjects with cognitive complaints reported by self or study partner (worried comparison subjects), 88 as "probable" mild cognitive impairment patients,[29,30] 106 as "possible" mild cognitive impairment patients (no complaint of cognitive decline, but with objective impairment on one or more primary neuropsychological variables), 55 as probable Alzheimer's disease patients, and 50 as possible Alzheimer's disease patients. 31 Participants diagnosed as cognitive comparison subjects (with or without complaints) were included if they had a Clinical Dementia Rating of 0,[32] an MMSE score ≥26,[33] and if they were not impaired on any primary neuropsychological test variable (i.e., no scores fell more than 1.5 standard deviations below normative means). Exclusion criteria included dominant hand hemiparesis or other central or peripheral motor impairments or visual acuity impairment that would preclude clock drawing test completion. The current study utilized the participants' most recent registry visit data.

Procedures

Trained psychometricians administered the clock drawing test in a standard way to include command (i.e., "I want you to draw the face of a clock, putting in all the numbers where they should go, and set the hands at 10 after 11") and copy conditions.[27] Participants were allowed to make corrections and make attempts to draw the clock a maximum of two times. Only the command condition data were used for the current study.

For the purpose of this study, 25 command clocks were randomly selected from each of the six diagnostic strata described above, resulting in the inclusion of 150 clocks from 150 different subjects with 50 clocks equally divided among each of the three primary diagnostic groups (i.e., comparison, mild cognitive impairment, Alzheimer's disease). The clocks were rated independently by four board-certified neurologists and a neurology nurse practitioner, all of whom specialize in dementia. Raters were blinded to participant diagnostic and demographic information. Ratings were made on a binary (normal/abnormal) and an ordinal scale (0–10 rating, where 0 signified no impairment and 10 signified complete impairment). The ordinal ratings were examined for potential outliers, and when widely discordant interexaminer ratings (score differences >5; arbitrarily selected) were found, these clocks were rerated by the original raters. The rerated clock drawing test scores were used in the analyses.

Statistical Analyses

All analyses were performed in SPSS 16 (Chicago) or SAS 9.1 (Cary, NC) software. A one-way analysis of variance (ANOVA) assessed for between-group differences in age, education, MMSE, and Geriatric Depression Scale scores. Significant findings were followed-up with Tukey-Kramer post-hoc tests to determine specific group differences. Sex and racial differences among the groups were analyzed using the chi-square test of independence.

Estimates of rater agreement were calculated for both the binary and ordinal rating scales. For the binary ratings, kappa statistics were computed.[34–36] Because there were five raters, the multiple agreement function in SAS was used for kappa calculations for the binary ratings. For the ordinal ratings, Kendall's intraclass correlation coefficient (ICC) of concordance was computed using the intraclass correlation functions in SAS, and Spearman rank correlations were generated between individual raters. Agreement for ordinal ratings were evaluated by calculating the absolute score differences between raters.

To examine the diagnostic utility (based on the three primary diagnostic groups) of the clinicians' ratings, sensitivity, specificity, and positive likelihood ratios were calculated for both the dichotomous and ordinal ratings. For the dichotomous ratings, two methods were used to summarize the five raters' data. First, we calculated individual sensitivity and specificity statistics for each rater and then created an average for all five raters. We also created a single summary rating of "impaired" versus "intact" based on whether the majority of the raters rated the clock as impaired or intact. For the ordinal ratings, an average was calculated to summarize the five clinicians' ratings, then sensitivities and specificities were calculated for this average rating against the true diagnostic classification. The ordinal ratings were used to examine the cutoff score that optimized sensitivity and specificity.

RESULTS

Details of demographics and clinical characteristics are shown in Table 1. There were significant between-group differences for age, female gender, education, and MMSE. There were no significant between-group differences in the Geriatric Depression Scale score.

The clinicians' interrater reliability was "almost perfect"[34–37] for the ordinal (ICC=0.92) system and "substantial" [37] for the dichotomous system (k=0.85). The absolute difference scores between ordinal ratings are presented in Table 2. The five clinicians' ratings did not differ by more than three ordinal scale units for 69% of the clocks. An absolute score difference of 5 or less was observed in 91% of the clocks, and only 5 clocks (3%) had an absolute difference of more than seven units on the ordinal scale. Examples of participant clocks that were rated similarly and dissimilarly are presented in Figure 1. Spearman rank correlations between each of the individual raters ranged from 0.64 to 0.82 for the ordinal scale (Table 3).

Based on the five clinicians' average dichotomous rating, the clinicians differentiated comparison and Alzheimer's disease participants with a sensitivity of 0.75 and a specificity of 0.81. In comparison, the dichotomous rating based solely on the majority of raters had a sensitivity of 0.84 and specificity of 0.84. In differentiating comparison subjects from mild cognitive impairment participants, the average sensitivity of the five raters' dichotomous classifications was 0.47 and the average specificity was 0.81. Using a majority for the calculation of sensitivity and specificity resulted in values of 0.50 and 0.84, respectively (Table 4).

The sensitivities, specificities, and positive likelihood ratios for several cutoff scores on the ordinal scale are presented in Table 5. For three of the four comparisons (i.e., Alzheimer's disease versus comparison, mild cognitive impairment versus comparison, and Alzheimer's disease + mild cognitive impairment versus comparison), a cutoff score of two or greater resulted in the maximization of sensitivity and specificity for differentiating diagnostic groups. For differentiating Alzheimer's disease from mild cognitive impairment, a cutoff score of four or greater was the rating that maximized sensitivity and specificity. As can be expected, higher ordinal ratings (i.e., more impaired clocks) were associated with a greater likelihood of being diagnosed with either Alzheimer's disease or mild cognitive impairment but at the expense of sensitivity.

DISCUSSION

The current study sought to investigate the interrater reliability of qualitative clock drawing test ratings made by five dementia clinicians at Boston University Medical Center. The clinicians were reliable clock drawing test raters using both dichotomous (impaired versus intact) and ordinal (0–10 impairment scale) ratings. The interrater reliability for the dichotomous system achieved a kappa of 0.85 and the ordinal rating resulted in an intraclass

correlation coefficient of 0.92. These statistics represent excellent interrater reliability values and are comparable to those obtained in our recent work comparing several widely used quantitative clock drawing test scoring systems.[27] The current findings demonstrate that in the absence of objective scoring methods, the clock drawing test can be rated reliably across a cognitive severity spectrum by clinicians who specialize in dementia.

Despite these excellent reliability values, there were several individual instances in which clinicians' ratings were widely disparate. As seen in Table 2, ratings of nine clocks (6%) differed by six or more points on the ordinal scale after rerating eliminated errors. There are multiple factors that may explain why the clinicians applied disparate ratings, including spatial configuration, participant self-corrected errors, and shape of the clock face as exemplified in Figure 1. The discrepancy among the raters highlights the difficulty that clinicians face when scoring clocks subjectively.

The present study also examined the accuracy of clinician-rated clock drawing test in differentiating among cognitively normal, mild cognitive impairment, and Alzheimer's disease diagnostic categories. Despite the substantial[37] overall agreement between raters, the results demonstrate that the accuracy with which qualitative ratings can differentiate diagnostic group membership was less robust. Although Alzheimer's disease patients and comparison subjects could be differentiated with a relatively high degree of accuracy, the ratings were considerably less useful when making the distinction between a diagnosis of mild cognitive impairment and comparison (less sensitive) or Alzheimer's disease and mild cognitive impairment (less specific). Therefore, while the clock drawing test may be a good screening instrument for Alzheimer's disease, it may not be a sensitive instrument for screening mild cognitive impairment, especially if clinicians use a dichotomous rating. When screening for mild cognitive impairment, the presence of an abnormal clock drawing test in isolation (based on subjective clinician rating) may result in a large number of false positive or false negative errors. For the mild cognitive impairment diagnosis, the sensitivity and specificity was somewhat improved by using a subjective ordinal rating scale with three or more cutoff points as compared to the dichotomous scale. We therefore suggest using a 3-point subjective ordinal clock drawing test rating scale such as "normal," "suspicious," and "impaired" to improve the mild cognitive impairment diagnosis rather than the existing dichotomous system.

The clinicians who served as raters for the current study are specialists in diagnosing dementia, and work in a tertiary care clinical setting and research center. Therefore, these clinicians may represent a more reliable and diagnostically accurate group than nonspecialists in the community. Their expertise in dementia assessment may limit the extent to which the findings can be generalized to other settings and clinicians. Another limitation is that some clinicians were also members of the consensus team that formulated the original diagnostic impressions for our participant cohort. This overlap raises the possibility that the clinicians may not have been completely blinded to diagnostic group membership for the clocks being reexamined, assuming that the clinicians remembered the clocks that were presented in prior consensus conference meetings. However, this overlap would have only impacted the diagnostic utility statistics and not the interrater reliability, which was the primary focus of the current study. We excluded individuals with dementia other than Alzheimer's disease, visual impairment and non-English speakers, which may have increased the diagnostic utility statistics while limiting the generalizability of our study.

Although the clock drawing test has many advantages as a screening instrument in the assessment of patients with suspected dementia, it is often used qualitatively, or subjectively, in clinical settings. As such, the reliability of these qualitative ratings between clinicians is brought into question. This is the first study to investigate the concordance among clock drawing test ratings by dementia specialists. The current study results indicate that dementia

specialists can reliably rate clock drawing test performance using two different qualitative rating approaches. In contrast, the findings do not support the use of the clock drawing test as a standalone screening instrument, as the classification accuracy statistics presented suggest that in mild cognitive impairment, the clinician ratings may be susceptible to both false positive and false negative errors. However, the clinicians' ratings had excellent sensitivity and specificity for distinguishing healthy comparison from probable and possible mild Alzheimer's disease. Future studies should compare the reliability and diagnostic accuracy of qualitative methods to empirically validated quantitative scoring systems.

Acknowledgments

This research was supported by National Institute of Health grants P30-AG13846 (Boston University Alzheimer's Disease Core Center), M01-RR00533 (Boston University General Clinical Research Center), K24-AG027841 (RCG), K23-AG030962 (Paul B. Beeson Career Development Award in Aging; ALJ), The authors thank Sabrina Poon, Melissa Barrup, Laura Byerly, Sita Yermashetti, Pallavi Joshi, Mario Orozco, Amanda Gentile and Kristen Huber for their assistance with data and all the psychometricians, nurse specialists and clinicians at the BU Alzheimer's Disease Center and GCRC for administering the clock drawing task and providing clinical assessments. In particular, the authors thank the participants of the Boston University Alzheimer's Disease Center cohort.

References

1. Hirtz D, Thurman DJ, Gwinn-Hardy K, et al. How common are the "common" neurologic disorders? Neurology 2007;68:326–337. [PubMed: 17261678]
2. Brookmeyer R, Johnson E, Ziegler-Grahamm K, et al. Forecasting the global prevalence and burden of Alzheimer's disease. Alzheimers Dement 2007;3:S168.
3. Alzheimer's Association: Alzheimer's Disease Facts and Figures. Alzheimers Dement 2008;4:110. [PubMed: 18631956]
4. Sager MA, Hermann BP, La Rue A, et al. Screening for dementia in community-based memory clinics. WMJ 2006;105:25–29. [PubMed: 17163083]
5. Valcour VG, Masaki KH, Curb JD, et al. The detection of dementia in the primary care setting. Arch Intern Med 2000;160 2964–298.
6. Borson S, Scanlan JM, Watanabe J, et al. Improving identification of cognitive impairment in primary care. Int J Geriatr Psychiatry 2006;21:349–355. [PubMed: 16534774]
7. Chodosh J, Petitti DB, Elliott M, et al. Physician recognition of cognitive impairment: evaluating the need for improvement. J Am Geriatr Soc 2004;52:1051–1109. [PubMed: 15209641]
8. Mok W, Chow TW, Zheng L, et al. Clinicopathological concordance of dementia diagnoses by community versus tertiary care clinicians. Am J Alzheimers Dis Other Demen 2004;19:161–165. [PubMed: 15214202]
9. Eddy JR, Sriram S. Clock-drawing and telling time as diagnostic aids. Neurology 1977;27:595. [PubMed: 559272]
10. Cahn-Weiner DA, Williams K, Grace J, et al. Discrimination of dementia with Lewy bodies from Alzheimer's disease and Parkinson's disease using the clock drawing test. Cogn Behav Neurol 2003;16:85–92. [PubMed: 12799594]
11. Royall DR, Mulroy AR, Chiodo LK, et al. Clock drawing is sensitive to executive control: a comparison of six methods. J Gerontol B Psychol Sci Soc Sci 1999;54:328–333.
12. Juby A, Tench S, Baker V. The value of clock drawing in identifying executive cognitive dysfunction in people with a normal mini-mental state examination score. Cmaj 2002;167:859–864. [PubMed: 12406943]
13. Lavery LL, Starenchak SM, Flynn WB, et al. The clock drawing test is an independent predictor of incident use of 24-hour care in a retirement community. J Gerontol A Biol Sci Med Sci 2005;60:928–932. [PubMed: 16079220]
14. Parker C, Philp I. Screening for cognitive impairment among older people in black and minority ethnic groups. Age Ageing 2004;33:447–452. [PubMed: 15217776]

15. Ferrucci L, Cecchi F, Guralnik JM, et al. Does the clock drawing test predict cognitive decline in older persons independent of the mini-mental state examination? The Fine Study Group. Finland, Italy, The Netherlands Elderly. J Am Geriatr Soc 1996;44:1326–1331. [PubMed: 8909348]
16. Death J, Douglas A, Kenny RA. Comparison of clock drawing with mini mental state examination as a screening test in elderly acute hospital admissions. Postgrad Med J 1993;69:696–700. [PubMed: 8255833]
17. Shulman KI, Shedletsky R, Silver IL. The challenge of time: clock-drawing and cognitive function in the elderly. Int J Geriatric Psychiatry 1986;1:135–140.
18. Sunderland T, Hill JL, Mellow AM, et al. Clock drawing in Alzheimer's disease: a novel measure of dementia severity. J Am Geriatr Soc 1989;37:725–729. [PubMed: 2754157]
19. Mendez MF, Ala T, Underwood KL. Development of scoring criteria for the clock drawing task in Alzheimer's disease. J Am Geriatr Soc 1992;40:1095–1109. [PubMed: 1401692]
20. Wolf-Klein GP, Silverstone FA, Levy AP, et al. Screening for Alzheimer's disease by clock drawing. J Am Geriatr Soc 1989;37:730–734. [PubMed: 2754158]
21. Rouleau I, Salmon DP, Butters N, et al. Quantitative and qualitative analyses of clock drawings in Alzheimer's and Huntington's disease. Brain Cogn 1992;18:70–87. [PubMed: 1543577]
22. Tuokko H, Hadjistavropoulos T, Miller JA, et al. The clock test: a sensitive measure to differentiate normal elderly from those with Alzheimer's disease. J Am Geriatr Soc 1992;40:579–584. [PubMed: 1587974]
23. Lam LC, Chiu HF, Ng KO, et al. 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.
24. Scanlan JM, Brush M, Quijano C, Borson S. Comparing clock tests for dementia screening: naive judgments vs formal systems–what is optimal? Int J Geriatr Psychiatry 2002;17:14–21. [PubMed: 11802225]
25. Morris JC, Heyman A, Mohs RC, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part 1: clinical and neuropsychological assessment of Alzheimer's disease. Neurology 1989;39:1159–1165. [PubMed: 2771064]
26. Jefferson AL, Wong S, Bolen E, et al. Cognitive correlates of HVOT test performance differ between individuals with mild cognitive impairment and normal controls. Arch Clin Neuropsychol 2006;21:405–412. [PubMed: 16893623]
27. Hubbard EJ, Santini V, Blankevoort CG, et al. Clock drawing performance in cognitively normal elderly. Arch Clin Neuropsychol 2008;23:295–327. [PubMed: 18243644]
28. Ashendorf L, Jefferson AL, O'Connor MK, et al. Trail making test errors in normal aging, mild cognitive impairment, and dementia. Arch Clin Neuropsychol 2008;23:129–137. [PubMed: 18178372]
29. Winblad B, Palmer K, Kivipelto M, et al. Mild cognitive impairment: beyond controversies, toward a consensus: report of the international working group on mild cognitive impairment. J Internal Med 2004;256:240–246. [PubMed: 15324367]
30. Petersen RC, Smith GE, Waring SC, et al. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999;56:303–308. [PubMed: 10190820]
31. McKhann GM, Drachman D, Folstein MF, et al. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group. Neurol 1984;34:939–944.
32. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurol 1993;43:2412–2414.
33. 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. [PubMed: 1202204]
34. Fleiss JL, Slakter MJ, Fischman SL, et al. Inter-examiner reliability in caries trials. J Dent Res 1979;58:604–609. [PubMed: 283091]
35. Fleiss, JL. Statistical Methods for Rates and Proportions. New York: John Wiley & Sons; 2003.
36. Kendall, MG. Rank correlation methods. London: Griffin; 1963.
37. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–174. [PubMed: 843571]

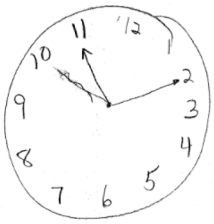
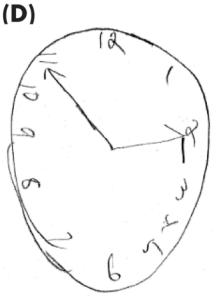
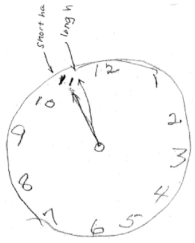
	Clocks Rated Similarly			Clocks Rated Dissimilarly		
Rater	Consensus Diagnosis	Abnormal	Rating	Consensus Diagnosis	Abnormal	Rating
1	(A)	N	1	(B)	Y	6
2		N	1		Y	2
3		N	1		N	0
4		N	0		N	1
5	Control	N	0	Control	Y	5
1	(C)	N	1	(D)	Y	8
2		N	0		Y	2
3		N	0		Y	6
4		N	0		N	2
5	MCI	N	0	MCI	N	0
1	(E)	Y	9	(F)	Y	9
2		Y	8		Y	4
3		Y	7		Y	8
4		Y	6		Y	3
5	AD	Y	8	AD	Y	4

FIGURE 1.
Concordant and Discordant Clinician Clock Drawing Test Ratings

J Neuropsychiatry Clin Neurosci. Author manuscript; available in PMC 2010 September 14.

TABLE 1

Participant Demographic Characteristics

	Comparison (n=50)		Mild Cognitive Impairment (n=50)		Alzheimer's Disease (n=50)		Analysis			
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**F**	**df**	χ^2	**Group Differences**
Age (years)[a]	72.0	9.3	72.2	10.1	79.0	9.0	8.76*	2, 147		Comparison=MCI<Alzheimer's Disease
Education (years)[a]	16.6	2.7	14.3	3.1	14.3	3.6	9.52*	2, 147		Comparison>MCI=Alzheimer's Disease
MMSE[a]	29.4	0.73	28.4	1.6	21.1	6.0	76.94*	2, 147		Comparison=MCI>Alzheimer's Disease
GDS[a]	3.9	4.3	4.1	4.0	5.4	3.7	1.85	2, 136		Comparison=MCI=Alzheimer's Disease
	n		**n**							
Gender (female)[b]	36		27		26			2	5.02	Comparison>MCI=Alzheimer's Disease
Ethnicity (Caucasian)[b]	43		30		40			2	9.97*	Comparison=Alzheimer's Disease<MCI

MCI=mild cognitive impairment; MMSE=Mini-Mental State Examination; GDS=Geriatric Depression Scale.

[a] Analysis of variance (degrees of freedom) with Tukey-Kramer post-hoc adjustments for multiple comparisons.

[b] Chi square test (degrees of freedom)

* $p<0.01$

TABLE 2

Absolute Difference of Scores for Individual Clock Ratings by Dementia Specialists

Difference in Rating	Frequency	%	Cumulative Frequency	Cumulative %
1	25	16.7	25	16.7
2	53	35.3	78	52.0
3	26	17.3	104	69.3
4	21	14.0	125	83.3
5	11	7.3	136	90.7
6	9	6.0	145	96.7
7	3	2.0	148	98.7
8	2	1.3	150	100.0

TABLE 3

Spearman Correlations for Clock Rating (0 –10) Between Clinicians

Rater	1	2	3	4
2	0.80*			
3	0.82*	0.79*		
4	0.77*	0.73*	0.79*	
5	0.78*	0.74*	0.66*	0.64*

*p<0.0001

TABLE 4

Sensitivity and Specificity of Dichotomous Ratings of Impairment

Clinicians Rating	Alzheimer's Disease versus Comparison		MCI versus Comparison		Alzheimer's Disease versus MCI		MCI + Alzheimer's Disease versus Comparison	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Average[a]	0.75	0.81	0.47	0.81	0.75	0.53	0.61	0.81
Majority[b]	0.84	0.84	0.50	0.84	0.84	0.5	0.67	0.84

MCI=mild cognitive impairment; N=50 each for comparison, MCI, and Alzheimer's disease group

[a]Average values across the five raters

[b]An abnormal rating was given if three or more raters agreed that the clock was abnormal.

TABLE 5

Sensitivity and Specificity for Various Cutoffs on the Ordinal Rating Scale (0 –10)

	Alzheimer's Disease versus Comparison			MCI versus Comparison			Alzheimer's Disease versus MCI			MCI + Alzheimer's Disease versus Comparison		
Cutoff*	**Sensitivity**	**Specificity**	**PLR**	**Sensitivity**	**Specificity**	**PLR**	**Sensitivity**	**Specificity**	**PLR**	**Sensitivity**	**Specificity**	**PLR**
≥**1**	0.98	0.46	1.8	0.82	0.46	1.5	0.98	0.18	1.2	0.90	0.46	1.7
≥**2**	0.84	0.76	3.5	0.58	0.76	2.4	0.84	0.42	1.4	0.71	0.76	3.0
≥**3**	0.66	0.90	6.6	0.40	0.90	4.0	0.66	0.60	1.7	0.53	0.90	5.3
≥**4**	0.54	0.94	9.0	0.26	0.94	4.3	0.54	0.74	2.1	0.40	0.94	6.7

MCI=mild cognitive impairment; PLR=positive likelihood ratio; n=50 each for comparison, MCI, and Alzheimer's disease groups.

*A higher score represents greater impairment.