



Neurología Argentina

www.elsevier.es/neurolarg



Original

El test del reloj: reproducibilidad, consistencia interna y variables predictivas de la prueba del reloj utilizando el método de puntuación de Cacho. Análisis de 985 relojes

Manuel Facundo Latini^{a,*}, Diana Scharovsky^b, Alejandra Glaser^c, Ruth Brugger^d, Juan Pablo Zorrilla^a, Leandro Sousa^a, Marina Sánchez^a, Alberto Alemán^a, Lucas Martín Romano^a y Sergio Eduardo Gonorazky^a

^aServicio de Neurología, Hospital privado de Comunidad, Mar del Plata, Argentina

^bJefa del Servicio de Atención a la Tercera Edad, Servicio de Neurología, Hospital privado de Comunidad, Mar del Plata, Argentina

^cServicio de Salud Mental, Servicio de Neurología, Hospital privado de Comunidad, Mar del Plata, Argentina

^dServicio de Atención a la Tercera Edad, Servicio de Neurología, Hospital privado de Comunidad, Mar del Plata, Argentina

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 30 de noviembre de 2010

Aceptado el 30 de diciembre de 2010

Palabras clave:

Test del reloj

Reproducibilidad

Pruebas de cribado

Demencia

Trastornos de memoria

R E S U M E N

Introducción: La prueba del reloj (PR) es una prueba de evaluación de cribado para deterioro cognitivo ampliamente difundida en la práctica clínica. Existen varios métodos de puntuación. En nuestro medio utilizamos el desarrollado por Cacho et al en 1999. El objetivo del trabajo es evaluar la reproducibilidad interevaluadores y la consistencia interna de la PR a la orden de pacientes que son derivados para evaluación de deterioro cognitivo utilizando el método de puntuación de Cacho, y analizar variables predictivas del resultado.

Material y métodos: Dos evaluadores entrenados ciegos entre sí y de otras evaluaciones y diagnósticos efectuados, valoraron los dibujos de reloj de 985 pacientes mediante el método de Cacho. Con los resultados y los datos de las historias clínicas se confeccionó una base de datos. Se utilizó el coeficiente kappa ponderado para evaluar la reproducibilidad interevaluadores y alfa de Cronbach para la consistencia interna. Se analizó la relación entre la puntuación del reloj con el sexo, la edad y los años de educación.

Resultados: La mediana del resultado de la PR de ambos evaluadores fue de 7. Se obtuvo un coeficiente kappa ponderado elevado (0,87 [95% IC 0,80-0,93]). La consistencia interna resultó regular en las pruebas realizadas por ambos observadores. La edad y la educación tendrían una correlación estadísticamente significativa pero de regular y escasa magnitud, negativa y positiva respectivamente.

Conclusiones: La reproducibilidad interevaluador fue similar a la de otros métodos de puntuación de la PR. La consistencia interna es menor que la observada por el autor original. La edad y los años de educación se correlacionan con los resultados de la prueba.

© 2010 Sociedad Neurológica Argentina. Publicado por Elsevier España, S.L.

Todos los derechos reservados.

*Autor para correspondencia.

Correo electrónico: Latinifacundo@gmail.com; latini_facundo@yahoo.com.ar (M.F. Latini).

The clock test: Reproducibility, consistency and predictors of the clock test using the scoring method of Cacho. Analysis of 985 clocks

A B S T R A C T

Keywords:

Clock drawing test
 Reproducibility of results
 Mass screening
 Dementia
 Memory disorders

Introduction: Clock drawing test (CDT) is a screening test for cognitive impairment widely used in clinical practice. There are several scoring methods. We use the one developed by Cacho et al in 1999. The objective was to evaluate the reproducibility and internal consistency of the CDT of patients referred for evaluation of cognitive impairment using the score method and to analyze predictors variables of the result.

Methods: Two trained raters blind to each other and other assessments and diagnostics, valued clock drawings of 985 patients by Cacho's method. With the results and data from hospital records a database was made. We used the weighted kappa coefficient to evaluate inter-rater reproducibility and Cronbach's alpha for internal consistency. We analyzed the relationship between the CDT results with gender, age and education.

Results: The median result of CDT of both evaluators was 7. We obtained a high weighted kappa coefficient (0.87 [95% CI 0.80 to 0.93]). Internal consistency was regular for both observers. Age and education have a statistically significant correlation but regular and small in magnitude, negative and positive respectively.

Conclusions: The reproducibility was similar to other scoring method. The internal consistency is lower than that observed by the original author. Age and years of education are correlated with test results.

© 2010 Sociedad Neurológica Argentina. Published by Elsevier España, S.L.
 All rights reserved.

Introducción

La prueba del reloj (PR) es ampliamente usada en la clínica diaria como técnica de cribado para evaluar el déficit cognitivo conjuntamente con otros test como la mini prueba del estado mental (MMSE) de Folstein o la escala de Blessed de concentración, atención y memoria^{1,2}. La MMSE está ampliamente difundida pero es un test verbal y evalúa fundamentalmente la memoria. Se sabe también que la MMSE puede ser deficiente para detectar deterioro cognitivo leve (DCL) o disfunción ejecutiva, por lo que sería insuficiente como única técnica de cribado¹⁻⁴.

La PR es un método rápido, de fácil aplicación en la práctica clínica incluso por parte de médicos en el primer nivel. Permite evaluar la memoria, la comprensión verbal, las facultades visuoespaciales, la planificación, el pensamiento abstracto y la concentración¹⁻³.

Se han desarrollado varios métodos de puntuación en diferentes países y con diferentes pacientes; como por ejemplo: Mendez⁵, Sunderland^{1,6}, Roleau^{1,7}, etc., pudiéndose evaluar a la copia y a la orden.

En nuestro medio utilizamos el desarrollado por Cacho et al en 1999⁸. Se desarrolló a partir de los criterios de Sunderland⁶ con la normas de la escala revisada por Roleau⁷ en 1992. En el estudio de Cacho se seleccionaron pacientes con diagnóstico clínico de enfermedad de Alzheimer mediante DSM-IV y NINCDS-ADRDA estadiados como demencia leve acorde a la escala CDR, apareados con sujetos normales.

La PR a la orden se puntúa en una escala del 0 al 10 puntos, evaluándose el dibujo de la esfera (0-2 puntos), las agujas (0-4

puntos) y los números (0-4 puntos). Para un punto de corte 6 mostró un 92,8% de sensibilidad y un 93,4% de especificidad⁸.

En el trabajo original no se evalúa la reproducibilidad entre observadores sino que se realiza un análisis de varianza de medidas repetidas destinada a probar hipótesis alternativas de diferencia entre observadores. Se sabe que este método, útil para rechazar hipótesis de nula diferencia, no es útil para probar la ausencia de ella, pues basta una muestra pequeña para obtener un resultado estadísticamente no significativo (error beta). Para probar la reproducibilidad entre observadores deben utilizarse pruebas como la de kappa o el coeficiente de correlación intraclass⁹.

El objetivo primario del trabajo es obtener la reproducibilidad entre las puntuaciones de reloj de pacientes que fueron derivados para evaluación de deterioro cognitivo, calificados por el método de Cacho, por dos evaluadores independientes ciegos entre sí y respecto a otras evaluaciones y diagnósticos de dichos pacientes.

Por otro lado, como en la evaluación de la consistencia interna mediante la técnica alfa de Cronbach el autor expresa que lo realiza estudiando "la correlación de cada ítem con la suma de los mismos", considerando más adecuado que la correlación interna debe efectuarse correlacionando los distintos ítems entre sí (y no con el total), consideramos como objetivo secundario la evaluación de este aspecto con nuestros datos.

El objetivo terciario es evaluar si variables como la edad, el sexo y los años de educación se correlacionan con la puntuación de la PR.

Material y métodos

El estudio se llevó a cabo en el SATE (Servicio de Atención a la Tercera Edad) del Hospital Privado de Comunidad, centro que se dedica a la evaluación, la contención y el tratamiento de pacientes con trastornos de memoria y en el que se trabaja con médicos geriatras, residentes de neurología, fonoaudiólogos, terapeutas ocupacionales y kinesiólogos. Los pacientes son derivados por médicos de cabecera, médicos clínicos o neurólogos para su evaluación. Los pacientes realizan una evaluación clínica, neurológica y neuropsicológica de cuyo consenso se obtiene una estadificación y el diagnóstico final.

Retrospectivamente se tomaron 985 evaluaciones de pacientes realizadas en los últimos 10 años. Personal administrativo realizó fotocopias de los dibujos de la PR a la orden de las respectivas historias (fig. 1), colocándole sólo el número de historia. Un grupo de investigadores recopiló de las historias clínicas los datos correspondientes a la edad en el momento de la evaluación, el sexo, los años de educación y las puntuaciones de escala de deterioro global (GDS), el índice clínico de demencia (CDR) y la MMSE, y la estadificación final, volcándolos en una base de datos.

Dos evaluadores (E1 y E2, con entrenamiento previo), ciegos entre sí y de los restantes resultados, re-puntaron por el método de Cacho los relojes fotocopados de las historias clínicas. Una vez finalizado el proceso, se volcó la información en la base de datos mencionada. Se realizó estadística descriptiva. Se obtuvo el coeficiente kappa ponderado a partir de los valores totales de cada prueba para analizar la concordancia entre los dos observadores. Se analizó la consistencia interna mediante el coeficiente alfa de Cronbach para cada evaluador por separado, correlacionando los tres ítems (esfera, agujas y números). Se analizó también la consistencia interna de la prueba en ambos evaluadores por separado con el subgrupo de pacientes con CDR=1 (a efectos de comparar los resultados con los publicados por Cacho).

Se obtuvo el promedio de los totales de ambos observadores. Se compararon los totales de las pruebas de ambos sexos mediante la prueba U de Mann-Whitney. Se correlacionaron los totales promediados de ambos evaluadores con las edades y con los años de educación mediante la prueba de Spearman. Se realizó un análisis multivariable mediante regresión lineal múltiple, considerando variables independientes la edad y los años de educación y el total promediado de la PR como

dependiente (se incluyeron sólo estas dos variables, pues ambas tenían una $p < 0,20$ en el bivariable).

Finalmente se correlacionó (Spearman) el promedio de las puntuaciones del reloj de ambos evaluadores con los resultados del CDR, la GDS y la MMSE y con la estadificación final. Para esto último se asignó un valor a cada uno de los diferentes estadios en orden creciente (a mayor severidad, mayor número). A pesar que originalmente los evaluadores que realizaron estas últimas pruebas no estaban cegados del resultado de la PR original, lo que puede constituir un sesgo, decidimos incluir estos datos en este trabajo debido a que no encontramos publicada en la bibliografía trabajos sobre correlación de dichas pruebas con PR puntuadas mediante el método de Cacho.

Se consideró estadísticamente significativa una $p_{2\alpha} < 0,05$.

Se utilizó el programa estadístico StatsDirect (versión 2.7.2).

Resultados

Los datos de los 985 pacientes se presentan en tabla 1.

La concordancia más allá del azar entre los dos evaluadores resultó elevada ($>0,80$). La consistencia interna de las evaluaciones fue regular en ambos evaluadores, con una mínima diferencia entre ambos (0,57 y 0,49, respectivamente) (tabla 2).

Tabla 1 – Datos del total de pacientes

Número de relojes	985 (947 pacientes)	
Sexo	68%	Mujeres
Edad	77 ^a	(88-71) ^b
Educación (años)	7 ^a	(12-5) ^b
GDS	3 ^a	(4-3) ^b
CDR	0,5 ^a	(1-0,5) ^b
MMSE	24 ^a	(27-20) ^b
Reloj evaluador 1	7 ^a	(8-5) ^b
Reloj evaluador 2	7 ^a	(8-4,5) ^b
Diagnósticos	n	Puntuación
Normales	10	8,3 ^a (10-6,5) ^b
Normales-DCAE	4	8 ^a (9-7,6) ^b
DCAE	40	8,6 ^a (9,75-7) ^b
DCAE-DCL	32	8 ^a (9,1-7) ^b
DCL	348	7,75 ^a (8,75-6,25) ^b
DCL-DML	101	7,5 ^a (8-5,5) ^b
DCL-DL	11	8 ^a (8-7,5) ^b
DML	231	5,75 (7,5-3,75) ^b
DML-DL	17	5 ^a (4,5-3,75) ^b
DL	149	6,5 ^a (8,25-4,5) ^b
DL-DM	2	5,5 ^a (9,5-1,5) ^b
DM	8	2,75 ^a (4,25-2) ^b
Sin diagnóstico	32	5,5 ^a (7,65-3) ^b

^aMediana.

^bRango intercuartil.

CDR: índice clínico de demencia; DCAE: declinación cognitiva asociada a la edad; DCL: deterioro cognitivo leve; DL: demencia leve; DM: demencia moderada-severa; DML: demencia muy leve; GDS: escala de deterioro global; MMSE: mini prueba del estado mental.

Figura 1 – Ejemplos de dibujos de prueba de reloj a la orden.

La consistencia interna de los pacientes del subgrupo con CDR=1 (217 pacientes) fue de 0,49 y 0,61 en cada evaluador (límite inferior del intervalo de confianza del 95% de 0,39 y 0,53, respectivamente).

No se encontraron diferencias estadísticamente significativa entre ambos sexos ($p_{2\alpha}=0,41$).

La edad y los años de educación tuvieron una correlación estadísticamente significativa ($p_{2\alpha}<0,001$) pero de regular y escasa magnitud, negativa y positiva respectivamente (Rho = -0,3 y 0,22) (tabla 3).

La PR promediada entre los dos observadores tuvo una moderada correlación positiva con la MMSE de Folstein. La correlación con la estadificación CDR y GDS fue negativa, de baja y moderada magnitud, respectivamente. Finalmente, también se observó moderada correlación negativa con la estadificación final (tabla 4).

Tabla 2 – Concordancia entre los dos evaluadores y consistencia interna

Kappa ponderado	0,87	95% IC, 0,80-0,93	$p < 0,0001$
Alfa de Cronbach E1	0,49	Límite inferior 95% IC = 0,44	
Alfa de Cronbach E2	0,57	Límite inferior 95% IC = 0,53	
IC: intervalo de confianza.			

Tabla 3 – Correlación y regresión lineal múltiple entre resultado de reloj y variables predictivas

Variable	Spearman	Multivariado
Edad	Rho = -0,3 ($p<0,001$ [95% IC, -0,35 a -0,24])	$r = -0,12$ ($p=0,0002$)
Años de educación	Rho = 0,22 ($p<0,001$ [95% IC, 0,16 a 0,18])	$r = 0,14$ ($p<0,0001$)

Tabla 4 – Correlación de Spearman entre resultados de reloj; test de evaluación de demencia y estadificación final

MMSE	0,53 (95% IC 0,48 a 0,57) ^a
CDR	-0,36 (95% IC -0,41 a -0,3) ^a
GDS	-0,4 (95% IC -0,49 a -0,39) ^a
Estadificación final	-0,49 (95% IC -0,5 a -0,44) ^a

^a $p<0,0001$

CDR: índice clínico de demencia; GDS: escala de deterioro global; MMSE: mini prueba del estado mental.

Discusión

En los últimos años hemos asistido a un auge en la utilización de la PR en la práctica clínica diaria. En el trabajo original de Cacho et al⁸ se estudiaron 56 pacientes con diagnóstico de enfermedad de Alzheimer. En nuestro trabajo se tomaron las evaluaciones de pacientes derivados por deterioro cognitivo independientemente de la etiología, con el fin de evaluar la PR en este tipo de población. En el trabajo original se evaluó la concordancia entre tres evaluadores mediante un análisis de varianza⁹. En nuestro análisis obtuvimos un coeficiente kappa ponderado elevado.

En un trabajo realizado en 2010 por Nair et al¹⁰, en donde se analizó la reproducibilidad interevaluadores utilizando un método cualitativo, obtuvieron un coeficiente kappa de 0,85. En un estudio de 2008 a cargo de Hubbard et al¹¹ se evaluó la reproducibilidad mediante coeficiente de correlación intraclass por diferentes métodos de puntuación, obteniéndose los siguientes resultados: Freund 0,80¹; Mendez 0,95^{1,5}; Cahn 0,91^{1,12}. En otro estudio de 2010 realizado por Aprahamian et al¹³ se obtuvo un coeficiente de correlación intraclass de 0,86 para el método de Shulman^{1,14} y de 0,90 para el método de Sunderland^{1,6}. En nuestro estudio el coeficiente kappa fue de 0,87, resultado consistente con una reproducibilidad similar al de otros métodos de puntuación de reloj.

Las consistencias internas que obtuvimos resultaron más bajas que las del trabajo original, aun analizando a pacientes con el mismo grado de demencia. Entendemos que ello se debe a una diferencia de metodología. En el estudio original se encontró un coeficiente alfa de entre 0,71 y 0,93 entre los diferentes ítems de la PR a la orden. En el estudio realizado por Lin et al¹⁵ se obtuvo un coeficiente alfa de 0,93.

En cuanto a variables predictivas como la edad y el nivel de educación, los trabajos publicados muestran resultados contradictorios^{1,8,16}. En nuestro análisis el sexo no fue una variable predictiva estadísticamente significativa. La edad y los años de educación fueron variables predictivas estadísticamente significativas pero de regular y escasa magnitud, respectivamente.

Comparado con el CDR y la GDS, la PR mostró baja y moderada correlación negativa. En un estudio llevado a cabo en 2002 por Powlisht et al¹⁷, la PR evaluada por diferentes métodos (que no incluía a Cacho) mostró un correlación con el CDR entre -0,69 y -0,74 entre pacientes con enfermedad de Alzheimer, mostrando además un pobre rendimiento en pacientes con demencia muy leve. Una revisión de 2010 a cargo de Ehreke et al¹⁸ señala que la PR no es una herramienta útil en el cribado de pacientes con DCL. En nuestro trabajo hay una gran cantidad de pacientes con demencia muy leve (DML) y DCL, lo que explicaría la baja correlación entre la PR y el CDR^{1,17-22}. Los estudios de Shulman et al^{1,14} (1993) y Sunderland et al^{1,6} (1989) muestran un regular nivel de correlación entre la PR y la GDS. La ausencia de cegamiento en el momento de la evaluación con la MMSE y la estadificación con CDR y GDS con la PR nos impide afirmar que los análisis realizados carecen de sesgo.

Si bien el método de Cacho es un método de puntuación validado en español y de fácil aplicación es nuestro medio, con una alta reproducibilidad, nos preocupa la regular consistencia interna observada por nosotros.

B I B L I O G R A F Í A

1. Pinto E, Peters R. Literature review of the clock drawing test as a tool for cognitive screening. *Dement Geriatr Cogn Disord.* 2009;27:201-13.
2. Kremer J. Clock drawing test in dementia: a critical review. *Neurol Arg.* 2002;27:223-7.
3. 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;167:859-64.
4. Cosentino S, Jefferson A, Chute DL, Kaplan E, Libon DJ. Clock drawing errors in dementia: neuropsychological and neuroanatomical considerations. *Cogn Behav Neurol.* 2004;17:74-84.
5. Mendez MF, Ala T, Underwood KL. Development of scoring criteria for the clock drawing task in Alzheimer's disease. *J Am Geriatr Soc.* 1992;40:1095-9.
6. Sunderland T, Hill JL, Mellow AM, Lawlor BA, Gundersheimer J, Newhouse PA, et al. Clock drawing in Alzheimer's disease, a novel measure of dementia severity. *J Am Geriatr Soc.* 1989;37:725-9.
7. 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.
8. Cacho J, García-García R, Arcaya J, Vicente JL, Lantada N. Una propuesta de aplicación y puntuación del test del reloj en la enfermedad de Alzheimer. *Rev Neurol.* 1999;28:648-55.
9. Gifford DR, Cummings JL. Evaluating dementia screening tests: methodologic standards to rate their performance. *Neurology.* 1999;52:224-7.
10. Nair AK, Gavett BE, Damman M, Dekker W, Green RC, Mandel A, et al. Clock drawing test ratings by dementia specialists: interrater reliability and diagnostic accuracy. *J Neuropsychiatry Clin Neurosci.* 2010;22:85-92.
11. Hubbard EJ, Santini V, Blankevoort CG, Volkens KM, Barrup MS, Byerly L, et al. Clock drawing performance in cognitively normal elderly. *Arch Clin Neuropsychol.* 2008;23:295-327.
12. Cahn DA, Salmon DP, Monsch AU, Butter 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;11:529-39.
13. Aprahamian I, Martinelli JE, Neri AL, Yassuda MS. The accuracy of the Clock Drawing Test compared to that of standard screening tests for Alzheimer's disease: results from a study of Brazilian elderly with heterogeneous educational backgrounds. *Int Psychogeriatr.* 2010;22:64-71.
14. Shulman KI, Gold DP, Cohen CA, Zuccherro CA. Clock-drawing and dementia in the community: a longitudinal study. *Int J Geriatr Psychiatry.* 1993;8:487-96.
15. Lin KN, Wang PN, Chen C, Chiu YH, Kuo CC, Chuang YY, et al. The three-item clock-drawing test: a simplified screening test for Alzheimer's disease. *Eur Neurol.* 2003;49:53-8.
16. Lourenço RA, Ribeiro-Filho ST, Moreira Ide F, Paradela EM, Miranda AS. The Clock Drawing Test: performance among elderly with low educational level. *Rev Bras Psiquiatr.* 2008;30:309-15.
17. Powlishta KK, Von Dras DD, Stanford A, Carr DB, Tsering C, Miller JP, et al. The clock drawing test is a poor screen for very mild dementia. *Neurology.* 2002;59:898-903.
18. Ehreke L, Luppá M, König HH, Riedel-Heller SG. Is the Clock Drawing Test a screening tool for the diagnosis of mild cognitive impairment? A systematic review. *Int Psychogeriatr.* 2010;22:56-63.
19. 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;53:353-7.
20. 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;19:119-27.
21. Pinto E, Peters R. Predictive value of the Clock Drawing Test. *Dement Geriatr Cogn Disord.* 2008;26:351-5.
22. Lim WS, Chong MS, Sahadevan S. Utility of the clinical dementia rating in Asian populations. *Clin Med Res.* 2007;5:61-70.