



ORIGINAL

Validez discriminante y concordancia interobservador de 2 métodos de puntuación del test del reloj



Teodoro del Ser^{a,*}, Belén Frades^a, Meritxell Valentí-Soler^a, María Ascensión Zea-Sevilla^a, Elizabeth Valeriano-Lorenzo^b y Cristóbal Carnero-Pardo^b

^a Unidad de Investigación Enfermedad de Alzheimer, Fundación CIEN. Instituto de Salud Carlos III, Madrid, España

^b FIDYAN Neurocenter, Granada, España

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 29 de abril de 2023

Aceptado el 23 de julio de 2023

On-line el 4 de septiembre de 2023

Palabras clave:

Test del reloj

Evaluación cognitiva

Diagnóstico

Deterioro cognitivo

RESUMEN

Objetivos: Comparar la validez discriminante y la fiabilidad interobservador de los 2 métodos de corrección del test del reloj más usados en España.

Metodología: Se han evaluado 2 colecciones de dibujos del reloj obtenidos en un contexto clínico (116 casos; 56,8% mujeres, edad media $73,1 \pm 7,7$ años) y en una cohorte de voluntarios (2.039 dibujos de 579 sujetos; 59,5% mujeres, edad media $78,3 \pm 3,8$ años). Todos los sujetos fueron clasificados como sin deterioro cognitivo (DC⁻) o con deterioro cognitivo (DC⁺) tras una extensa evaluación clínica y neuropsicológica. Evaluadores expertos han valorado estos dibujos de forma independiente y sin conocimiento del diagnóstico con los métodos de Sunderland y Solomon estandarizados en español por Cacho (rango: 0 a 10) y del Ser (rango: 0 a 7), respectivamente. Se ha calculado la validez discriminante de cada método mediante el área bajo la curva ROC (aROC) en las 2 muestras, y la fiabilidad interobservador mediante el coeficiente de correlación intraclass (CCI) y el coeficiente kappa en la muestra clínica que fue valorada por los 2 evaluadores.

Resultados: No hay diferencias significativas en la validez discriminante de los métodos de Sunderland y Solomon en ninguna de las muestras (clínica: aROC: 0,73 [IC 95%: 0,64-0,81] y 0,77 [IC 95%: 0,69-0,85], respectivamente, $p=0,19$; voluntarios: aROC: 0,69 [IC 95%: 0,67-0,71] y 0,72 [IC 95%: 0,69-0,73], respectivamente, $p=0,08$). Los puntos de corte ≤ 8 y ≤ 5 clasifican correctamente al 71 y 73% de la muestra clínica y al 82 y 84% de la muestra de voluntarios, respectivamente. Los 2 métodos tienen una buena concordancia en la muestra clínica (Sunderland: CCI: 0,90 [IC 95%: 0,81-0,93]; kappa: 0,76 [IC 95%: 0,70-0,83]; Solomon: 0,92 [IC 95%: 0,88-0,95] y 0,77 [IC 95%: 0,71-0,83], respectivamente), algo mayor en el segundo, aunque las diferencias no son significativas.

Conclusiones: La validez discriminante y la fiabilidad interobservador de estos 2 métodos de corrección del test del reloj son similares. El método de Solomon, más breve y sencillo, puede ser más aconsejable en términos pragmáticos.

© 2023 SEGG. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

Discriminant validity and inter-rater concordance of two scoring systems for the clock test

A B S T R A C T

Objective: To compare the discriminant validity and inter-rater reliability of the two scoring systems for the Clock test that are most used in Spain.

Methodology: Two collections of clock drawings obtained in a clinical context (116 cases; 56.8% women, mean age 73.1 ± 7.7 years) and in a cohort of volunteers (2039 drawings of 579 subjects; 59.5% women, mean age 78.3 ± 3.8 years) have been assessed. All subjects were classified as cognitively normal (CN) or cognitively impaired (CI) after extensive clinical and neuropsychological evaluation. Expert raters have evaluated these drawings independently and without knowledge of the diagnosis using the Sunderland and Solomon systems standardized in Spanish by Cacho (range 0 to 10) and del Ser (range 0 to 7)

Keywords:

Clock test

Cognitive assessment

Diagnosis

Cognitive impairment

* Autor para correspondencia.

Correo electrónico: teodorodelser@gmail.com (T. del Ser).

respectively. The discriminant validity of each method was calculated in the two samples using the area under the ROC curve (aROC), and the inter-rater reliability was calculated in the clinical sample, that was assessed by the two evaluators, using the intraclass correlation coefficient (ICC) and the *kappa* coefficient. **Results:** There are no significant differences in the discriminant validity of the Sunderland and Solomon systems in any of the samples (clinical: aROC 0.73 [CI95%: 0.64-0.81] and 0.77 [CI95%: 0.69-0.85] respectively, $P=.19$; volunteers: aROC 0.69 [CI95%: 0.67-0.71] and 0.72 [CI95%: 0.69-0.73] respectively, $P=.08$). The cut-off points ≤ 8 and ≤ 5 correctly classify 71% and 73% of the clinical sample and 82% and 84% of the volunteer sample, respectively. Both systems have good agreement in the clinical sample (Sunderland: ICC 0.90 [CI95%: 0.81-0.93], *kappa* 0.76 [CI95%: 0.70-0.83]; Solomon: 0.92 [CI95%: 0.88-0.95] and 0.77 [CI95%: 0.71-0.83] respectively), somewhat higher in the second, although the differences are not significant.

Conclusions: The discriminant validity and inter-observer reliability of these two Clock Test correction systems are similar. Solomon's method, shorter and simpler, may be more advisable in pragmatic terms.

© 2023 SEGG. Published by Elsevier España, S.L.U. All rights reserved.

Introducción

El test del reloj (TdR) fue diseñado inicialmente para evaluar la función visoespacial¹, pero se ha convertido en uno de los test cognitivos de cribado más utilizados, tanto de forma individual² como integrado en otros instrumentos como el Mini-Cog³, el MoCA⁴, el test de los 7 min⁵, el Rapid Cognitive Screen⁶ o el Addenbrooke Cognitive Examination III⁷.

El TdR es breve, simple, y tiene buena fiabilidad interobservador^{8,9}. También tiene buena sensibilidad para detectar el deterioro cognitivo (DC)¹⁰, ya que exige el concurso de múltiples funciones cognitivas como atención, comprensión verbal, memoria visual, habilidad visoespacial, programación motora, abstracción y respuesta inhibitoria^{1,2}. No está exento de inconvenientes, especialmente en la evaluación de personas mayores, analfabetas o escasamente escolarizadas, ya que requiere una destreza grafomotora y está influido por la edad y el nivel educativo^{8,10-12}. Tiene un buen rendimiento en la discriminación de los casos con demencia frente a controles, pero mucho menor en la diferenciación de los casos con deterioro cognitivo ligero¹³.

Se han propuesto numerosos métodos de puntuación para el TdR, tanto cuantitativos como cualitativos^{9,14}, sin que ninguno haya demostrado una utilidad sustancialmente mayor^{9,14-17}. En España se han validado y estandarizado 2 métodos de puntuación del TdR^{2,18-20}. Ambos métodos son cuantitativos, pero dan puntuaciones diferentes a los elementos del dibujo. El primero de ellos^{2,18} utiliza unos criterios similares a los de Sunderland et al.²¹ y Rouleau et al.²² y una escala de 0 a 10. El segundo^{19,20}, incorporado por Solomon en el test de los 7 min^{5,20}, es una versión abreviada del método de Freedman et al.²³ y utiliza una escala de 0 a 7. No existe por el momento ninguna comparación de las cualidades psicométricas de estos 2 métodos de corrección que permita hacer una elección razonada entre ellos.

El objetivo de este estudio es evaluar un gran número de dibujos del TdR mediante estos 2 métodos de puntuación y calcular la validez discriminante y la concordancia interobservador de cada uno de ellos para comparar su utilidad clínica en el cribado de deterioro cognitivo.

Metodología

Participantes

Se han evaluado 2 colecciones de dibujos del TdR. La primera de ellas incluye 116 dibujos obtenidos en 52 pacientes de la consulta de CC y 64 sujetos de un estudio clínico caso control²⁴. Los sujetos fueron reclutados durante la evaluación clínica de un posible DC. La segunda incluye 2.039 dibujos realizados por 579 voluntarios, mayores de 69 años, reclutados en el proyecto Vallecás, un estudio

longitudinal de los determinantes del deterioro cognitivo en una cohorte de personas mayores²⁵.

Procedimientos

En todos los sujetos se realizó un examen clínico y neurológico, y una extensa evaluación neuropsicológica. Tras ello fueron clasificados como sin deterioro cognitivo (DC⁻) o con DC⁺ según criterios estándares de deterioro cognitivo ligero²⁶ y demencia²⁷ en los que la puntuación en el TdR no fue tenida en cuenta. En la mayoría de los sujetos el DC era de probable origen neurodegenerativo, pero no se han utilizado marcadores diagnósticos y no se ha considerado la etiología en el análisis.

Todos los individuos cuya información clínica y dibujos del reloj han sido analizados en este estudio habían proporcionado el consentimiento informado para usar sus datos anonimizados en investigación clínica. El proyecto Vallecás fue aprobado por el Comité Ético del Instituto de Salud Carlos III.

Dos profesionales expertos examinaron los dibujos de forma independiente y sin conocimiento del diagnóstico. La colección clínica fue evaluada y puntuada por ambos evaluadores con los métodos de Sunderland y de Solomon; la colección de voluntarios fue evaluada con un solo método por cada uno de ellos. En todos los casos se siguieron los criterios de estandarización en español^{18,19}.

El método de Sunderland valora en una escala de 0 a 10 la esfera (0 a 2 puntos), la presencia y secuencia de los números (0 a 4 puntos) y la presencia y ubicación de las manecillas (0 a 4 puntos). El método de Solomon valora en una escala de 0 a 7 la presencia y disposición de los números (0 a 3 puntos) y de las manecillas (0 a 4 puntos).

Variables

Se han recogido para este estudio las variables demográficas edad, sexo y nivel educativo (> estudios primarios, estudios primarios hasta los 10 años y < estudios primarios), así como las puntuaciones de cada dibujo obtenidas con cada uno de los 2 métodos.

Análisis de datos

Se han comparado las variables demográficas y las puntuaciones de los sujetos con DC⁺ y DC⁻ mediante la *t* de Student para las variables continuas y el χ^2 para las variables categoriales. Se ha estimado la relación entre las puntuaciones obtenidas en el TdR con los 2 métodos mediante el estadístico *r* de Pearson. La relación de las variables demográficas con las puntuaciones en el TdR se ha examinado mediante análisis de regresión lineal múltiple.

La validez discriminante de cada uno de los métodos de puntuación del TdR se ha calculado mediante el área bajo la curva

Tabla 1
Datos demográficos y puntuaciones de los dibujos

	Muestra de voluntarios ^a (n = 2.039)			Muestra clínica ^a (n = 116)		
	No DC	DC	Valor de p	No DC	DC	Valor de p
N.º	1.808 (88,7%)	231 (11,3%)		42 (36,2%)	74 (63,8%)	
Edad (años)	78,1 ± 3,7 [69-89]	79,3 ± 4,3 [71-92]	**	73,4 ± 1,1 [61-86]	72,9 ± 0,9 [49-88]	ns
Sexo (mujer)	1.076 (59,5%)	151 (65,4%)	ns	22 (52,4%)	42 (56,8%)	ns
Estudios:			**			*
<Primarios	262 (14,5%)	40 (17,3%)		14 (33,3%)	18 (24,3%)	
Primarios	527 (29,1%)	97 (42,0%)		8 (19,0%)	33 (44,6%)	
>Primarios	1.019 (56,4%)	94 (40,7%)		20 (47,6%)	23 (31,1%)	
Sunderland	9,6 ± 0,9 [1-10]	8,4 ± 2,1 [1-10]	**	8,0 ± 2,0 [2-10]	6,0 ± 2,6 [1-10]	**
Solomon	6,6 ± 0,8 [1-7]	5,4 ± 1,7 [0-7]	**	5,7 ± 1,5 [1-7]	3,8 ± 2,0 [0-7]	**

DC: deterioro cognitivo; No DC: no deterioro cognitivo; ns: no significativo.

Los datos son: media ± desviación típica [rango] para las variables cuantitativas y n.º de sujetos (porcentaje) para las cualitativas. En la muestra de voluntarios las cifras están referidas a los dibujos del reloj analizados. Se han hecho las comparaciones estadísticas mediante la t de Student para las variables continuas y el χ^2 para las variables categoriales.

^a En la muestra clínica cada sujeto aporta un dibujo del reloj. En la muestra de voluntarios 579 sujetos aportan 2.039 relojes dibujados en un número variable de evaluaciones anuales sucesivas (fig. 1).

* 0,01.

** <0,001.

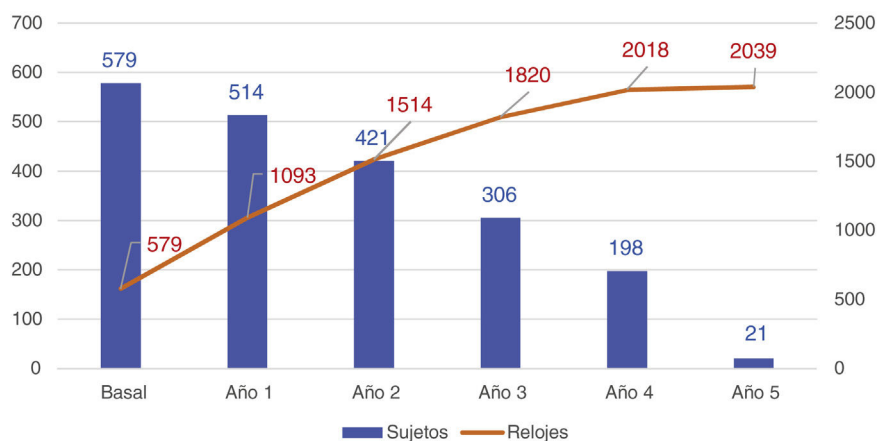


Figura 1. Sujetos, relojes y n.º de evaluaciones en la muestra de voluntarios. Las barras indican el número de voluntarios que realizan cada evaluación y el dibujo del reloj. La línea representa el número de relojes que se acumulan con las evaluaciones sucesivas.

ROC (aROC). La comparación entre las curvas se ha realizado por el método de Hanley y McNeil²⁸. En ambas muestras se han calculado además, la sensibilidad, la especificidad, el cociente de probabilidad positivo y el porcentaje de clasificaciones correctas. Para estos cálculos se han utilizado los puntos de corte habitualmente aceptados en la clínica: ≤ 8 en el método Sunderland y ≤ 5 en el método Solomon, pero se han explorado también otros puntos de corte. La fiabilidad interobservador se ha analizado mediante los coeficientes de correlación intraclass (CCI) y el coeficiente kappa en los 116 casos de la muestra clínica que fueron valorados por los 2 evaluadores.

Los cálculos se han llevado a cabo con el programa SPSS® 19²⁹; en los resultados se ha considerado un error alfa del 5% bilateral o un intervalo de confianza del 95%.

Resultados

La mayoría de los sujetos estudiados son mayores de 60 años. La edad es ligera y significativamente mayor en los sujetos con DC⁺ solo en la muestra de voluntarios; no hay diferencias por sexo, y el nivel educativo de los sujetos con DC⁺ es significativamente inferior en ambas muestras (tabla 1).

Los sujetos de la muestra clínica fueron evaluados una sola vez y diagnosticados de DC⁻ (42 casos) o DC⁺ (74 casos; 38 deterioro cognitivo ligero y 36 demencia); cada uno aporta un dibujo del reloj.

Los 579 voluntarios del proyecto Vallecas fueron evaluados en un número variable de ocasiones sucesivas con una periodicidad anual y aportan un total de 2.039 dibujos del reloj; el número de relojes correspondiente a cada evaluación anual se representa en la figura 1. El número de sujetos con DC⁺ y DC⁻ cambia en cada evaluación anual de seguimiento; la muestra completa incluye 1.808 relojes dibujados por sujetos con DC⁻ (88,7%) y 231 por sujetos con DC⁺ (11,3%; 202 con deterioro cognitivo ligero y 29 con demencia) en la evaluación correspondiente.

Las puntuaciones obtenidas con ambos métodos son significativamente inferiores en los dibujos de los sujetos con DC⁺ (tabla 1), y muestran entre sí una alta y significativa correlación (r de Pearson [IC 95%]: 0,76 [0,74-0,78] en la muestra de voluntarios; 0,84 [0,78-0,89] en la muestra clínica).

Hay una estrecha y significativa relación positiva de la educación y negativa del diagnóstico de DC con las puntuaciones en el TdR en ambas muestras y métodos; la edad y el sexo femenino solo muestran una relación negativa y significativa en la muestra de voluntarios que es más numerosa (tabla 2).

La validez discriminante de cada método se presenta mediante el aROC (tabla 3 y fig. 2); es mayor en ambas muestras con el método Solomon (muestra clínica: $0,77 \pm 0,04$; muestra de voluntarios: $0,72 \pm 0,02$; $p < 0,001$ en ambas muestras) que con el método de Sunderland (muestra clínica: $0,73 \pm 0,05$; muestra de voluntarios: $0,69 \pm 0,02$; $p < 0,001$ para ambas), pero las diferencias entre áreas (muestra clínica: $0,045 \pm 0,03$; $z = 1,18$; $p = 0,24$, muestra de

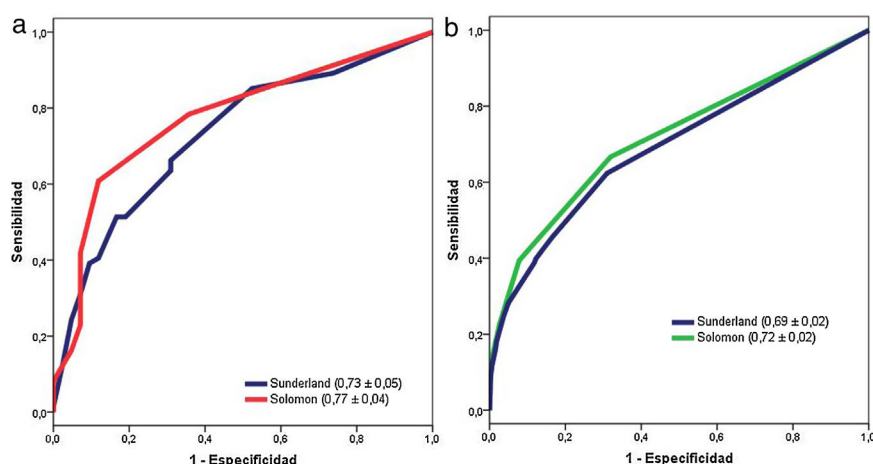


Figura 2. Curvas ROC obtenidas con cada método de evaluación. a) Muestra clínica; b) Muestra Vallecas.

Tabla 2

Relaciones entre las variables demográficas y las puntuaciones en el test del reloj

	Muestra de voluntarios (n = 2.039)		Muestra clínica (n = 116)	
	Puntuaciones en el test del reloj		Puntuaciones en el test del reloj	
	Sunderland	Solomon	Sunderland	Solomon
Edad	-0,028	-0,043*	-0,048	-0,050
Sexo femenino	-0,043*	-0,075**	-0,038	-0,076
Nivel educativo	0,221**	0,258**	0,345***	0,258**
Diagnóstico de DC	-0,297**	-0,334*	-0,361**	-0,312**

DC: deterioro cognitivo.

Los datos son los coeficientes estandarizados b de los análisis de regresión lineal múltiple sobre la variable dependiente «puntuación en el dibujo del test del reloj» con cada uno de los métodos en cada una de las 2 muestras.

* < 0,05.

** < 0,0001.

Tabla 3

Validez discriminante de los dos métodos de puntuación del TdR

	Muestra de voluntarios (n = 2.039)			Muestra clínica (n = 116)		
	aROC ± ee	IC 95%	Valor de p	aROC ± ee ^a	IC 95% ^b	Valor de p
Método Sunderland	0,69 ± 0,02	0,67-0,71		0,73 ± 0,05	0,64-0,81	
Método Solomon	0,72 ± 0,02	0,69-0,73		0,77 ± 0,04	0,69-0,85	
Diferencia entre áreas	0,025 ± 0,01		0,08	0,045 ± 0,03		0,19

aROC ± ee: área bajo la curva ROC ± error estándar; IC 95%: intervalo de confianza del 95%; TdR: test del reloj.

voluntarios: $0,025 \pm 0,01$; $z = 1,59$; $p = 0,11$) no alcanzan significación estadística.

La [tabla 4](#) presenta los principales parámetros métricos de ambos métodos con diferentes puntos de corte. Los puntos de corte ≤ 8 en el método Sunderland y ≤ 5 en el método Solomon, que son los que ofrecen mejores rendimientos diagnósticos en la muestra clínica, clasifican correctamente al 82% de la muestra de voluntarios y al 71% de la clínica con el primero, y al 84% de la muestra de voluntarios y al 73% de la clínica con el segundo ([tabla 4](#)). Los puntos de corte ≤ 6 en el método Sunderland y ≤ 4 en el método Solomon, que son los que ofrecen mejores rendimientos diagnósticos en la muestra de voluntarios por la menor prevalencias de DC, clasifican correctamente al 89% de la muestra de voluntarios y al 63% de la clínica con el primero, y al 89% de la muestra de voluntarios y al 71% de la clínica con el segundo ([tabla 4](#)).

Las diferencias en las puntuaciones obtenidas por los dos evaluadores independientes en la muestra clínica (Solomon: $0,1 \pm 0,8$ y Sunderland: $-0,04 \pm 1,1$; t para muestras apareadas 1,3 y $-0,4$, respectivamente) no son significativas. Los índices de fiabilidad interobservador son ligeramente mayores con el método Solomon (CCI: 0,92 [IC 95%: 0,89-0,95]; $kappa$: $0,77 \pm 0,03$

[IC 95%: 0,71-0,83]) que con el método Sunderland (CCI: 0,90 [IC 95%: 0,86-0,93]; $kappa$: $0,76 \pm 0,03$ [IC 95%: 0,70-0,83]), pero las diferencias no alcanzan significación estadística.

Discusión

Numerosos estudios han comparado varios métodos de puntuación del TdR en la detección del deterioro cognitivo y la mayoría concluyen que todos ellos recogen un núcleo común de actividad cognitiva y tienen cualidades métricas bastante similares^{15,17,30-32}. Por ello, la decisión sobre cuál utilizar depende de las preferencias y entrenamiento del evaluador o del tiempo y recursos disponibles. Nuestros datos son concordantes con estos estudios comparativos, pero al haber sido obtenidos en 2 muestras independientes, numerosas, y de procedencia y estructura diferentes, son especialmente robustos. Estos resultados indican que un método de puntuación más complejo no proporciona mayor validez discriminante ni concordancia entre evaluadores, aunque cabe esperar que aumente el tiempo y el coste de la evaluación¹⁵. En este estudio, el método de Solomon, que utiliza una escala más simple, ofrece una validez discriminante y una fiabilidad interobservador

Tabla 4

Parámetros métricos de los 2 métodos de puntuación del test del reloj

	PdC	Muestra de voluntarios (n = 2.039)				Muestra clínica (n = 116)			
		S	E	CP*	CC	S	E	CP*	CC
Método Sunderland	≤ 1	0,00	0,99	7,83	0,88	0,01	100	—	0,37
	≤ 2	0,01	0,99	15,65	0,88	0,12	0,98	5,11	0,43
	≤ 3	0,05	0,99	28,70	0,88	0,24	0,95	5,11	0,50
	≤ 4	0,08	0,99	29,74	0,89	0,31	0,93	4,35	0,53
	≤ 5	0,12	0,99	31,31	0,89	0,39	0,90	4,11	0,57
	≤ 6	0,16	0,98	9,91	0,89	0,51	0,83	3,08	0,63
	≤ 7	0,25	0,96	6,66	0,88	0,63	0,69	2,05	0,65
	≤ 8	0,39	0,88	3,28	0,82	0,85	0,48	1,63	0,71
	≤ 9	0,45	0,83	2,76	0,79	0,88	0,33	1,32	0,68
	≤ 10	1,00	0,00	1,00	0,11	1,00	0,00	1,00	0,64
Método Solomon	≤ 1	0,04	0,99	39,13	0,88	0,16	0,95	3,41	0,45
	≤ 2	0,09	0,99	52,18	0,89	0,23	0,93	3,22	0,48
	≤ 3	0,15	0,99	17,74	0,89	0,41	0,93	5,86	0,59
	≤ 4	0,23	0,97	8,47	0,89	0,61	0,88	5,11	0,71
	≤ 5	0,39	0,90	5,05	0,84	0,78	0,64	2,19	0,73
	≤ 6	0,67	0,68	2,09	0,68	0,86	0,40	1,45	0,70
	≤ 7	1,00	0,00	1,00	0,11	1,00	0,00	1,00	0,64

CC: clasificaciones correctas; CP*: cociente de probabilidad positivo; E: especificidad; PdC: punto de corte; S: sensibilidad.

ligera pero no significativamente superiores al método de Sunderland.

El rendimiento diagnóstico del TdR puede mejorarse combinando métodos de puntuación cuantitativos y cualitativos^{9,33}, o asociándolo a otras pruebas cognitivas^{13,34}. Recientemente se están aplicando nuevas tecnologías, como la utilización de un lápiz digital³⁵, o métodos de inteligencia artificial para mejorar la lectura y el tratamiento de los dibujos del TdR. Estos métodos ofrecen buenos rendimientos diagnósticos y pueden representar el futuro de la evaluación de este y otros test de cribado cognitivo^{36,37}.

Conclusiones

La validez discriminante y la fiabilidad interobservador de estos 2 métodos de corrección del TdR son similares. Pero nuestros datos y otros recogidos de la literatura indican que, para el cribado del deterioro cognitivo, tanto en el entorno clínico como poblacional, un método de puntuación del TdR más simple, como el de Solomon, ofrece los mismos resultados que el más complejo de Sunderland, por lo que parece más práctico y razonable su uso¹⁵⁻¹⁷.

Autorías

- Dr. Del Ser: concepto y diseño del estudio, interpretación de los datos, supervisión de los análisis estadísticos y redacción/revisión del artículo.
- B. Frades: adquisición y evaluación de los dibujos, interpretación de los datos y revisión del artículo.
- M. Valentí, M.A. Zea y E.L. Valeriano: adquisición de los dibujos, interpretación de los datos y revisión del artículo.
- Dr. Carnero-Pardo: concepto y diseño del estudio, realización de los análisis estadísticos, interpretación de los datos y redacción/revisión del artículo.

Financiación

La presente investigación no ha recibido ayudas específicas provenientes de agencias del sector público, sector comercial o entidades sin ánimo de lucro.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses. No tienen ningún empleo, actividades de consultoría, honorarios,

acciones, patentes o becas relacionadas con el contenido de esta investigación.

El Dr. T. del Ser dirigió hace años la estandarización para la población española mayor de 70 años de la versión de Solomon del TdR.

Cinco de los autores son empleados de la Unidad de Investigación Enfermedad de Alzheimer de la Fundación CIEN, que está financiada por el Instituto de Salud Carlos III y la Fundación Reina Sofía, y han trabajado en el proyecto Vallecas de donde proceden gran parte de los datos de este estudio. Los restantes datos proceden del centro FYDIAN regentado por el otro autor.

Agradecimientos

Agradecemos la colaboración de los voluntarios del Proyecto Vallecas y los pacientes de la clínica FYDIAN.

Bibliografía

1. Freedman M, Leach L, Kaplan E, Winocour G, Shulman K, Delis S. Clock drawing. A neuropsychological analysis. New York: Oxford University Press; 1994.
2. Cacho Gutiérrez J, García García R, García García-Patino R. El Test del Reloj. En: Carnero Pardo C, editor. Test cognitivos breves. Madrid: SEN Ediciones; 2015. p. 121-44.
3. Borson S, Scanlan J, Brush M, Vitaliano P, Dokmak A. The mini-cog: A cognitive 'vital signs' measure for dementia screening in multi-lingual elderly. Int J Geriatr Psychiatry. 2000;15:1021-7. [http://dx.doi.org/10.1002/1099-1166\(200011\)15:11<1021::aid-gps234>3.0.co;2-6](http://dx.doi.org/10.1002/1099-1166(200011)15:11<1021::aid-gps234>3.0.co;2-6).
4. Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc. 2005;53:695-9. <http://dx.doi.org/10.1111/j.1532-5415.2005.53221.x>.
5. Solomon PR, Hirschhoff A, Kelly B, Relin M, Brush M, DeVeaux RD, et al. A 7 minute neurocognitive screening battery highly sensitive to Alzheimer's disease. Arch Neurol. 1998;55:349-55. <http://dx.doi.org/10.1001/archneur.55.3.349>.
6. Malmstrom TK, Voss VB, Cruz-Oliver DM, Cummings-Vaughn LA, Tumosa N, Grossberg GT, et al. The Rapid Cognitive Screen (RCS): A Point-of-Care Screening for Dementia and Mild Cognitive Impairment. J Nutr Health Aging. 2015;19:741-4. <http://dx.doi.org/10.1007/s12603-015-0564-2>.
7. Bruno D, Schurmann Vignaga S. Addenbrooke's cognitive examination III in the diagnosis of dementia: A critical review. Neuropsychiatr Dis Treat. 2019 Feb 15;15:441-7. <http://dx.doi.org/10.2147/NDT.S151253>.
8. Mazancova AF, Nikolai T, Stepankova H, Kopecek M, Bezdicek O. The reliability of Clock Drawing Test scoring systems modeled on the normative data in healthy aging and nonamnestic Mild Cognitive Impairment. Assessment. 2017;24:945-57. <http://dx.doi.org/10.1177/1073191116632586>.
9. Martínez-Arán A, Solé P, Salamero M, de Azpiroz P, Tomás S, Marín R. El test del dibujo del reloj: métodos de evaluación cuantitativos y cualitativos. Rev Neurol. 1998;27:55-9.
10. Kim H, Chey J. Effects of education, literacy, and dementia on the Clock Drawing Test performance. J Int Neuropsychol Soc. 2010;16:1138-46. <http://dx.doi.org/10.1017/S1355617710000731>.

11. Maestri G, Nicotra A, Pomati S, Canevelli M, Pantoni L, Cova I. Cultural influence on clock drawing test: A systematic review. *J Int Neuropsychol Soc.* 2022;25:1–11, <http://dx.doi.org/10.1017/S1355617722000662>.
12. de Noronha ÍFC, Barreto SDS, Ortiz KZ. The influence of education on performance of adults on the Clock Drawing Test. *Dement Neuropsychol.* 2018;12:61–7, <http://dx.doi.org/10.1590/1980-57642018dn12-010009>.
13. Rubínová E, Nikolai T, Marková H, Siffelová H, Laczó J, Hort J, et al. Clock Drawing Test and the diagnosis of amnesic mild cognitive impairment: Can more detailed scoring systems do the work? *J Clin Exp Neuropsychol.* 2014;36:1076–83, <http://dx.doi.org/10.1080/13803395.2014.977233>.
14. Spenciere B, Alves H, Charchat-Fichman H. Scoring systems for the Clock Drawing Test: A historical review. *Dement Neuropsychol.* 2017;11:6–14, <http://dx.doi.org/10.1590/1980-57642016dn11-010003>.
15. Mainland BJ, Amodeo S, Shulman KI. Multiple clock drawing scoring systems: Simpler is better. *Int J Geriatr Psychiatry.* 2014;29:127–36, <http://dx.doi.org/10.1002/gps.3992>.
16. Kim S, Jahng S, Yu KH, Lee BC, Kang Y. Usefulness of the Clock Drawing Test as a cognitive screening instrument for Mild Cognitive Impairment and mild dementia: An evaluation using three scoring systems. *Dement Neurocogn Disord.* 2018;17:100–9, <http://dx.doi.org/10.12779/dnd.2018.17.3.100>.
17. Fuentes Pérez MA, de la Fuente Cadenas JA, Guerrero Desirré J, Castell Fríguls E. Test del reloj. ¿Existe un sistema de puntuación óptimo en atención primaria? *Aten Primaria.* 2008;40:476–7.
18. 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.
19. Del Ser T, Sánchez-Sánchez F, García de Yébenes MJ, Otero A, Zunzunegui MV, Munoz DG. Versión española del Test de los Siete Minutos. Datos normativos de una muestra poblacional de ancianos de más de 70 años. *Neurología.* 2004;19:344–58.
20. Del Ser Quijano T. El Test de los 7 Minutos. En: Carnero Pardo C, editor. *Test cognitivos breves*. Madrid: SEN Ediciones; 2015. p. 105–20.
21. 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, <http://dx.doi.org/10.1111/j.1532-5415.1989.tb02233.x>.
22. Rouleau I, Salmon DP, Butters N, Kennedy C, McGuiire K. Quantitative and qualitative analyses of clock drawings in Alzheimer's and Huntington's disease. *Brain Cogn.* 1992;18:70–87, [http://dx.doi.org/10.1016/0278-2626\(92\)90112-y](http://dx.doi.org/10.1016/0278-2626(92)90112-y).
23. Freedman M, Leach L, Kaplan E, Winocur G, Shulman KI, Dells D. *Clock drawing: A neuropsychological analysis*. New York: Oxford University Press Inc; 1994.
24. Carnero-Pardo C, Montoro-Ríos MT. Evaluación preliminar de un nuevo test de cribado de demencia (Eurotest). *Rev Neurol.* 2004;38:201–9.
25. Olazarán J, Valentí M, Frades B, Zea-Sevilla MA, Ávila-Villanueva M, Fernández-Blaquez MÁ, et al. The Vallecas Project: A cohort to identify early markers and mechanisms of Alzheimer's disease. *Frontiers in Aging Neuroscience.* 2015;7:181, <http://dx.doi.org/10.3389/fnagi.2015.00181>.
26. Albert MS, Dekosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging and Alzheimer's Association workgroup. *Alzheimer's Dement.* 2011;7:270–9, <http://dx.doi.org/10.1016/j.jalz.2011.03.008>.
27. American Psychiatric Association (APA). *DSM-IV-TR*. Barcelona: Masson; 2000.
28. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology.* 1983;148:839–43, <http://dx.doi.org/10.1148/radiology.148.3.6878708>.
29. IBM Corp. Released 2010. IBM SPSS Statistics for Windows, Version 19.0. Armonk, NY: IBM Corp.
30. South MB, Greve KW, Bianchini KJ, Adams D. Interrater reliability of three clock drawing test scoring systems. *Appl Neuropsychol.* 2001;8:174–9, <http://dx.doi.org/10.1207/S15324826AN0803.7>.
31. Van der Burg M, Bouwen A, Stessens J, Ylief M, Fontaine O, de Lepeleire J, et al. Scoring clock tests for dementia screening: A comparison of two scoring methods. *Int J Geriatr Psychiatry.* 2004;19:685–9, <http://dx.doi.org/10.1002/gps.1143>.
32. García-Caballero A, Recimil MJ, García-Lado I, Gayoso P, Cadarso-Suárez C, González-Hermida J, et al. ACE clock scoring: A comparison with eight standard correction methods in a population of low educational level. *J Geriatr Psychiatry Neurol.* 2006;19:216–9, <http://dx.doi.org/10.1177/0891988706292759>.
33. Dong F, Shao K, Guo S, Wang W, Yang Y, Zhao Z, et al. Clock-drawing test in vascular mild cognitive impairment: Validity of quantitative and qualitative analyses. *J Clin Exp Neuropsychol.* 2020;42:622–33, <http://dx.doi.org/10.1080/13803395.2020.1793104>.
34. Brodaty H, Moore CM. The Clock Drawing Test for dementia of the Alzheimer's type: A comparison of three scoring methods in a memory disorders clinic. *Int J Geriatr Psychiatry.* 1997;12:619–27.
35. Souillard-Mandar W, Penney D, Schaible B, Pascual-Leone A, Au R, Davis R. DCT-clock: Clinically-interpretable and automated artificial intelligence analysis of drawing behavior for capturing cognition. *Front Digit Health.* 2021;3:750661, <http://dx.doi.org/10.3389/fdgth.2021.750661>.
36. Amini S, Zhang L, Hao B, Gupta A, Song M, Karjadi C, et al. An artificial intelligence-assisted method for dementia detection using images from the Clock Drawing Test. *J Alzheimers Dis.* 2021;83:581–9, <http://dx.doi.org/10.3233/JAD-210299>.
37. Jiménez-Mesa C, Arco JE, Valentí-Soler M, Frades-Payo B, Zea-Sevilla MA, Ortiz A, et al. Using explainable artificial intelligence in the Clock Drawing Test to reveal the cognitive impairment pattern. *Int J Neural Systems.* 2023;33:2350015, <http://dx.doi.org/10.1142/S0129065723500156>.