

ORIGINAL

Rendimiento diagnóstico de los test cognitivos aplicados desde atención primaria. Concordancia y validez de los test de cribado

Teófilo Lorente Aznar^{a,b,*}, Francisco Javier Olivera Pueyo^c, Sergio Benabarre Ciria^b, Mariano Rodríguez Torrente^b, Beatriz Solans Aisa^d y Ana Carmen Giménez Baratech^e

^aCentro de Salud de Sabiñánigo, Huesca, España

^bUnidad Docente de Medicina Familiar y Comunitaria, Instituto Aragonés de Ciencias de la Salud (I+CS), Huesca, España

^cUnidad de Salud Mental “Pirineos”, Instituto Aragonés de Ciencias de la Salud (I+CS), Huesca, España

^dCentro de Salud Luis Vives I, Alcalá de Henares, Madrid, España

^eCentro de Salud de Berdún, Huesca, España

Recibido el 11 de mayo de 2009; aceptado el 7 de julio de 2009

Disponible en Internet el 11 de noviembre de 2009

PALABRAS CLAVE

Deterioro cognitivo;
Ancianos;
Fiabilidad y validez

Resumen

Objetivo: Estimar la prevalencia de deterioro cognitivo (DC) en ancianos de la comunidad. Evaluar el rendimiento diagnóstico de los principales test de cribado cognitivo usados en atención primaria: minisexamen cognoscitivo (MEC), test del reloj (TR), test de fluencia verbal (TFV) y test del informador (TIN). Calcular la concordancia entre ellos y su validez como pruebas de cribado.

Diseño: Estudio descriptivo transversal.

Participantes: Doscientos noventa y tres personas de ambos sexos, mayores de 64 años, residentes en la provincia de Huesca, seleccionados mediante muestreo aleatorizado sistemático. Marco muestral: base de datos de Tarjeta Sanitaria Individual. Ámbito provincial.

Mediciones: Variables sociodemográficas básicas, y exploración del DC con MEC, TR, TFV y TIN.

Resultados: En función del MEC y los diagnósticos previos, la prevalencia de DC es del 18,1% (IC del 95%: del 15,3 al 24,4%).

La concordancia del MEC con el resto de los test de cribado es moderada: con el TR kappa (κ)=0,40, con el de TFV: 0,33, con el TIN: 0,36.

La concordancia global del conjunto de los test cognitivos es moderada-baja: κ (Jackknife)=0,35 (IC del 95%: 0,269 a 0,444).

Considerando patrón de oro al MEC junto con el diagnóstico previo de DC, el TR presenta una sensibilidad del 78,3% y una especificidad del 76,9%; el TFV presenta el 80 y el 70%,

*Autor para correspondencia.

Correo electrónico: tlorenteaznar@gmail.com (T. Lorente Aznar).

respectivamente, y el TIN presenta el 76,7 y el 71,4%, respectivamente. El área bajo la curva ROC (*receiver operating characteristics*) del TFV es significativamente superior al resto. **Conclusión:** La concordancia entre los distintos test de cribado cognitivo es moderada. El TFV es el que mejor rendimiento diagnóstico presenta.

© 2009 Elsevier España, S.L. Todos los derechos reservados.

KEYWORDS

Cognitive impairment;
Elderly;
Reliability and validity

Diagnostic yield of cognitive tests applied in primary care. Consistency and validity of screening tests

Abstract

Objective: To estimate the prevalence of cognitive impairment in elderly people living in the community. To assess the diagnostic yield of the main cognitive screening tests in Primary Care: the Mini Mental State Examination (MMSE), the Clock Drawing Test (CDT), the Verbal Fluency Test (VFT) and the Informant Questionnaire test (IQ). We also calculated their correlation and validity.

Design: Descriptive study based on a personal interview by professionals using a standardised method to assess cognitive impairment.

Participants: A total of 293 men and women, who were older than 64 years old and residing in the province of Huesca. The sample was recruited by systematic random sampling from the database of personal health care card.

Main measurements We analysed social and demographic variables (age, educational level, marital status) and the assessment of cognitive impairment by means of MMSE, CDT, VF, IQ tests.

Results: The prevalence of cognitive impairment with MMSE was 18.1% (95% CI: 15.3–24.4%). The correlation between MMSE and the other cognitive screening tests was moderate: CDT kappa=0.40; VF kappa=0.33; IQ kappa=0.36. Using the MMSE as a gold standard and the prior diagnosis of CI, the sensitivity of CDT was 78.3% and the specificity was 76.9%, 80%; there was 70% sensitivity and 76.7% specificity for the VF; and 71.4% for IQ respectively. The area under the VF ROC curve was significantly higher than the other tests.

Conclusions: The agreement between cognitive impairment screening tests was moderate. The VF yielded better diagnostic accuracy.

© 2009 Elsevier España, S.L. All rights reserved.

Introducción

El deterioro cognitivo (DC), en sus diferentes manifestaciones, es altamente prevalente en atención primaria. Se estima que el DC leve (*mild cognitive impairment*) puede afectar en los países occidentales hasta al 20% de los mayores de 65 años¹, mientras que la demencia lo hace al 5–10% en este grupo de edad, y llegó al 30% en la década de 1980². Sin embargo, el interés del tema no reside únicamente en su elevada prevalencia, sino también en sus consecuencias, dado que ambos procesos son la principal causa de discapacidad en los mayores de 60 años³. Se ha demostrado que el DC leve presenta una elevada tendencia evolutiva hacia la demencia (entre un 12 y un 15% cada año)⁴, de manera que es especialmente importante reconocer precozmente este trastorno, con el objeto de aplicar de forma temprana programas de prevención secundaria que la retarden.

Por otra parte, es igualmente conocido que el DC se encuentra infradiagnosticado en la comunidad⁵. Se han desarrollado numerosos instrumentos para su cribado; el Grupo de Trabajo de Demencias de la semFYC ha definido algunos de éstos como útiles para su aplicación en atención primaria⁶, y se utilizan ampliamente tanto en la práctica clínica como en investigación⁷.

Sin embargo, se han observado prevalencias de DC variables en función de la prueba utilizada⁸, posiblemente en relación con las importantes diferencias de validez interna de los diferentes instrumentos más comúnmente utilizados: miniexamen cognoscitivo (MEC) de Lobo, test del reloj (TR), set-test de Isaacs, Short Portable Mental Status Questionnaire de Pfeiffer, test de los 7min, etc. Obviamente, esto tiene consecuencias en cuanto a la tarea de aproximación diagnóstica individual y de seguimientos posteriores, la organización de los servicios y la asignación de recursos.

El desarrollo y la validación de estos test se han realizado preferentemente en el ámbito de la atención especializada y, por tanto, con población seleccionada. Son escasos los estudios de ámbito comunitario, por lo que resulta difícil conocer si este hecho desempeña algún papel en la pobre correlación observada entre las diferentes pruebas. Se ha comprobado, además, que cuando se utilizan pruebas complementarias o a los resultados del test se asocia la opinión del cuidador, se incrementa la sensibilidad diagnóstica y aumenta la prevalencia de deterioro detectada⁹.

Se hace necesario, pues, conocer al detalle la sensibilidad de los diferentes test de cribado de DC, las pruebas e ítems que presentan una mayor rentabilidad diagnóstica y también los efectos de una posible asociación entre ellos y con un

test aplicado al informador/cuidador. Estos hechos son de especial importancia en el marco de la atención primaria, que es donde los test breves de cribado pueden mostrar su utilidad real, especialmente cuando profesionales de este ámbito los aplican, de manera que puedan elegirse los más adecuados, detectar el DC de forma temprana e implementar programas de prevención secundaria que retrasen el progresivo desarrollo de un cuadro de demencia o faciliten el tratamiento precoz de ésta una vez instaurada.

El presente estudio tiene por objetivos estimar la prevalencia de DC en pacientes de edad igual o superior a los 65 años que residen en la comunidad en el ámbito de la provincia de Huesca, observar las diferencias en las prevalencias obtenidas en función del instrumento aplicado, calcular el grado de concordancia entre los principales test de cribado cognoscitivo que se utilizan en atención primaria y determinar su validez de criterio utilizando como patrón de referencia la versión española del Mini Mental State: el MEC de Lobo.

Material y métodos

Tipo de estudio y selección de la población

Se ha realizado un estudio descriptivo de base poblacional, que incluye participantes de ambos性, de edad igual o superior a 65 años, domiciliados en la provincia de Huesca y seleccionados de forma aleatorizada mediante un muestreo sistemático, empleando como marco de sondeo la base de datos de Tarjeta Sanitaria Individual.

El tamaño de la muestra se ha calculado en base a la estimación de una concordancia medida en forma de índice de kappa (κ). En previsión de un resultado en torno al 0,5, con un desvío del 0,06 y una confianza del 95% resultaría un tamaño muestral de 267 participantes. Al considerar que se van a analizar subgrupos y objetivos secundarios, y que el equipo puede asumir un tamaño algo mayor, éste finalmente se fija en 324 participantes.

Se informó a los participantes en el estudio de las características de éste, y se solicitó su consentimiento escrito para participar.

Variables e instrumentos

Se recogieron variables generales epidemiológicas: sexo, edad, población de residencia, centro de salud, nivel de escolarización, etnia, existencia de ingresos hospitalarios en el último año, viudedad reciente, alteraciones sensoriales, relaciones sociales, existencia o no de aislamiento social (convivientes), mala salud autopercibida, insomnio y cambio reciente de domicilio.

Como variables de comorbilidad se registraron la presencia de depresión (escala abreviada de depresión geriátrica de Yesavage)¹⁰, ansiedad (subescala de ansiedad de la escala de ansiedad y depresión de Goldberg)¹¹, sintomatología psicótica (ítems relacionados con los trastornos del pensamiento y la sensopercepción del Geriatric Mental State)¹² y gravedad física (índice acumulativo de enfermedad)^{13,14}.

La posible existencia de DC se exploró mediante los siguientes test:

- MEC de Lobo¹⁵: se trata de la versión española del Mini-Mental State Examination de Folstein. Es, probablemente, el test más empleado en nuestro país y en el resto del mundo para el cribado del DC. Personal no médico puede aplicarlo en apenas 10 min; ha demostrado ampliamente su fiabilidad, validez y poder discriminativo, con una especificidad del 82% y una sensibilidad del 84,6%, que en pacientes geriátricos asciende. Es un instrumento de uso habitual en las consultas de atención primaria. Se ha utilizado la versión de 35 puntos con punto de corte en 24 para personas mayores de 65 años.
- TR¹⁶: es un test basado en el dibujo de un reloj por parte del paciente. De extraordinaria facilidad en cuanto a su realización, se ordena al paciente que dibuje una esfera de reloj con sus números y las manecillas que marquen las 11.10. La puntuación se establece sobre la base del cumplimiento de cada uno de los aspectos parciales de la orden. Su punto de corte se ha establecido en 6. Es útil por sí solo y también asociado a pruebas cortas de valoración del estado mental e incrementa la sensibilidad de éstas, por eso lo consideramos de gran utilidad en nuestro medio y en la práctica clínica.
- Test del informador (TIN)¹⁷: se trata de un test autocumplimentado por el informador, familiar o cuidador que acompaña al paciente en relación con los cambios en la memoria y la inteligencia detectados en los últimos 10 años y que colabora aportando o aclarando los datos del cuestionario general. Se ha utilizado la versión abreviada de 17 ítems, con el punto de corte en 57; por encima de éste se considera probable la existencia de DC.
- Test de fluencia verbal (TFV)¹⁸: se ha empleado una de las categorías del set-test de Isaacs, concretamente el subtest de evocación de animales en un minuto, y se ha utilizado un dintel variable en función del nivel de instrucción: 10 para escolaridad baja, 12 para el nivel medio y 13 para el alto. Por encima de estos valores se consideraba posible DC.

Recogida de datos

La recogida de datos se realizó por medio de entrevistas personales que llevaron a cabo médicos y enfermeros/as de atención primaria de cada uno de los centros de salud participantes, y un psicólogo clínico, todos ellos entrenados en la administración de test cognitivos. En los casos en los que se detectaron problemas de comunicación con los participantes, las entrevistas se realizaron con familiares o cuidadores fiables.

Se confeccionó un cuaderno de vaciado de datos que contenía las distintas preguntas y test diagnósticos. Se pilotó este cuaderno previamente para detectar problemas de fiabilidad o interpretación. Se realizaron reuniones periódicas con los médicos y personal de enfermería y psicología participantes con funciones de entrevistadores, con el objeto de detallar las características del estudio, sus fases, para unificar criterios y aclarar dudas.

Análisis estadístico

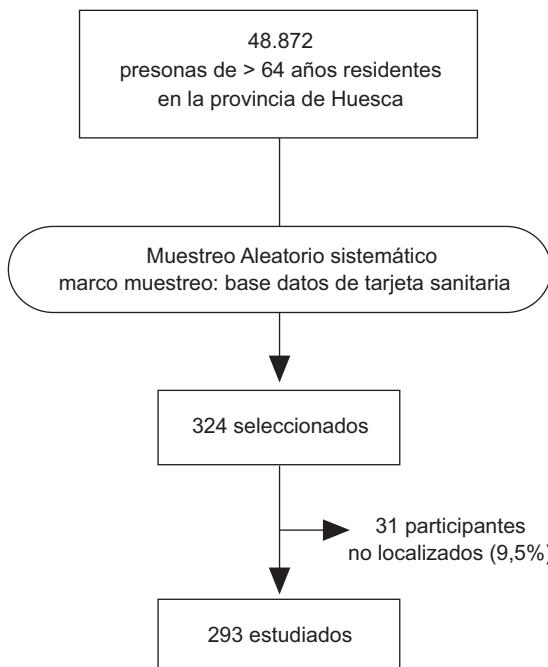
Se ha realizado un análisis descriptivo univariante que estudia la distribución de frecuencias de las distintas variables registradas y calcula los estadísticos de tendencia central y dispersión en el caso de los datos cuantitativos.

Las estimaciones de las prevalencias de DC se presentan con intervalos de confianza (IC) del 95%.

Para el estudio de concordancia entre los distintos test, se calculó el coeficiente de correlación intraclase, así como la concordancia global entre los resultados dicotómicos de los test y el índice κ resultante entre ellos^{19,20}.

La validez de criterio de los distintos test se evaluó contrastando sus resultados con los del MEC de Lobo y calculando para cada uno de ellos su sensibilidad, especificidad, razones de probabilidad, valores predictivos positivo y negativo, y construyendo las curvas ROC para establecer un punto de corte, comparando las áreas bajo las curvas ROC mediante el procedimiento de Handley y McNeil²¹.

Los datos se han tratado con las aplicaciones estadísticas SPSS® y Epidat®.



Esquema general del estudio. Estudio descriptivo transversal de una muestra aleatoria de habitantes de Huesca.

Resultados

De la muestra inicial de 324 participantes, se estudian finalmente 293 pacientes por no poder localizar a 31 de ellos (9,5%).

Entendiendo como casos de DC aquéllos detectados mediante el MEC durante el estudio, además de aquellos pacientes ya diagnosticados previamente (constancia en los antecedentes reflejados en la lista de problemas de su historia clínica), se encontraron 53 casos, y resultó una prevalencia del 18,1% (IC del 95%: 15,3-24,4%). Si se aceptan los resultados de los otros test como nuevos casos

Tabla 1 Concordancia del miniexamen cognoscitivo de Lobo (n=290) con el resto de los test

	TR n=273	TFV n=289	TIN n=140
Concordancia global, %	77,1	71,7	72,4
Kappa de Cohen, %	40,1	33,5	36,0

TFV: test de fluencia verbal; TIN: test del informador; TR: test del reloj.

de DC, resultaría una prevalencia del 19,5% según el TIN, del 30,4% según el TR y del 38,2% según el TFV.

En la tabla 1 pueden apreciarse la concordancia global entre los resultados del MEC y el resto de los test, así como los índices κ . La concordancia conjunta de los 3 test más empleados (MEC, TFV y TR) fue estadísticamente significativa, pero en valores moderados: κ (Jackknife)=0,36 (IC del 95%: 0,27-0,44; Z=10,31; p<0,001).

La concordancia entre el valor de los puntos totales del MEC con los de los otros test fue estadísticamente significativa en los 3 casos, y alcanzó los valores más elevados en el caso del TFV (coeficiente de correlación intraclase=0,66, IC del 95%: 0,59-0,72; F=4,86; p<0,001) (fig. 1).

En la tabla 2 se presentan los resultados del análisis de la validez de criterio de los test, y destacan los valores predictivos negativos altos, en torno al 94%, para todos ellos.

Las áreas bajo las curvas ROC son para el TFV de 0,886 (IC del 95%: 0,841-0,930), para el TR de 0,838 (0,778-0,898) y para el TIN de 0,776 (0,655-0,896). Al compararlas con el procedimiento de Handley y McNeill²¹ se observa que el área bajo la curva del TFV es significativamente mayor que la del TIN (Z=2,03, p<0,05) (fig. 2).

Discusión

Los instrumentos que muestran una mayor coincidencia respecto a las prevalencias obtenidas son el MEC y el TIN, con valores del 18,1 y el 19,1% respectivamente. Puede esto considerarse un resultado esperable, si tenemos presentes la sensibilidad y especificidad previamente publicadas y muy próximas para ambos (el 89,8 y el 83,9% para el MEC y el 86 y el 92%, respectivamente, para el TIN)^{15,22}. Sin embargo, nuestros datos no revelan para este último una sensibilidad y especificidad tan elevadas, probablemente debido a una cierta selección de la muestra utilizada, ya que en nuestro trabajo únicamente se han evaluado con el TIN aquellas personas que acudían a la evaluación acompañados por un cuidador, lo que, obviamente, no sucedía en todos los casos. Por otra parte, aunque hay autores que señalan que la discriminación del TIN es mayor que la del Mini Mental State Examination, al presentar la ventaja de no verse influenciado por el nivel educativo ni el deterioro funcional del paciente evaluado, no debe pasarse por alto su mediatisación por la fiabilidad del informante²³.

Los 2 instrumentos con los que se obtienen mayores prevalencias son el TFV y el TR. Una explicación de la elevada prevalencia de probable DC que ofrece el TFV, a

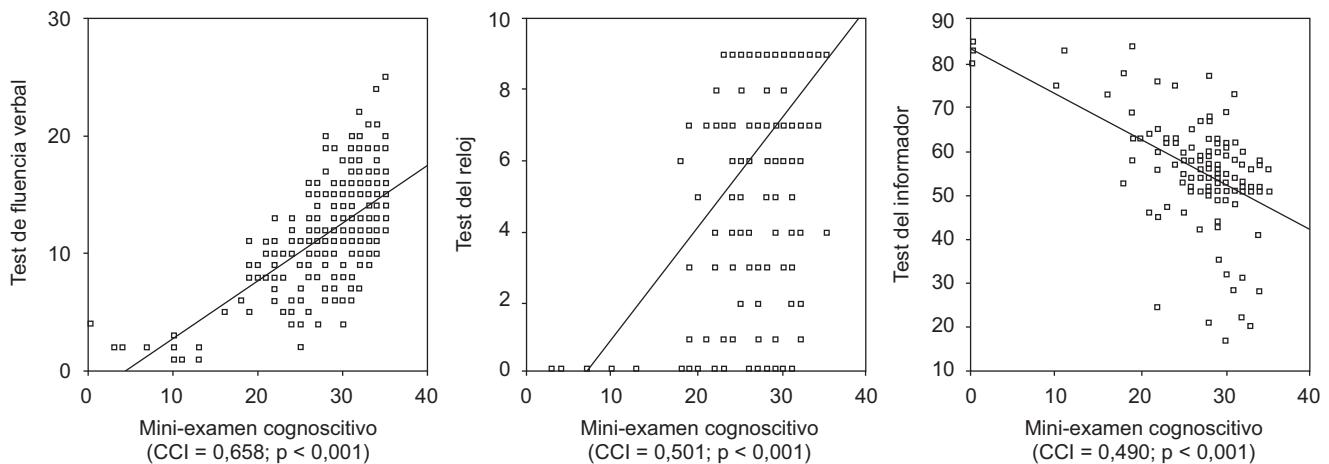


Figura 1 Correlación de los resultados del miniexamen cognoscitivo con el resto de los test de cribado.

Tabla 2 Validez de criterio de los test*

	TFV	TR	TIN
Pacientes correctamente diagnosticados, % (IC del 95%)	71,72 (66,1–76,7)	77,09 (71,5–81,8)	72,48 (64,4–79,3)
Sensibilidad, % (IC del 95%)	80,00 (65,8–89,5)	78,26 (63,2–88,5)	76,67 (57,3–89,3)
Especificidad, % (IC del 95%)	70,00 (63,7–75,6)	76,86 (70,7–82,0)	71,43 (62,3–79,1)
Valor predictivo positivo, % (IC del 95%)	35,71 (27,0–45,3)	40,45 (30,3–51,4)	40,35 (27,8–54,1)
Valor predictivo negativo, % (IC del 95%)	94,38 (89,6–97,1)	94,62 (90,0–97,2)	92,39 (84,4–96,6)
Razón de probabilidades positivo	2,67 (2,10–3,38)	3,38 (2,55–4,48)	2,68 (1,9–3,7)
Razón de probabilidades negativo	0,29 (0,16–0,5)	0,28 (0,16–0,49)	0,33 (0,17–0,63)

IC: intervalo de confianza; TFV: test de fluencia verbal; TIN: test del informador; TR: test del reloj.

Datos para una prevalencia de deterioro cognitivo del 18,1%.

* Los casos de enfermedad se definen por tratarse de pacientes previamente diagnosticados de deterioro cognitivo antes de iniciar el estudio, junto con aquellos que se diagnostican al presentar positividad en el miniexamen cognoscitivo de Lobo.

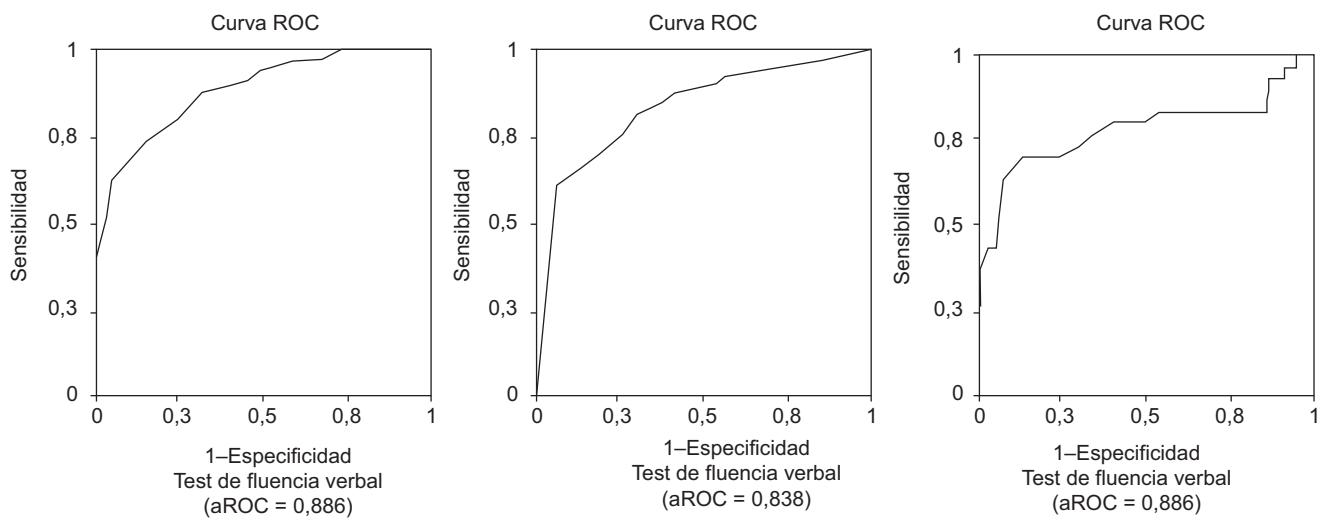


Figura 2 Curvas ROC (*receiver operating characteristics*) de los distintos test.

pesar de ser el menos sensible de los utilizados en el estudio según la bibliografía (79%)¹⁸, podría ser el bajo nivel educativo de la muestra utilizada (casi el 50% de los participantes presentaba un nivel de instrucción bajo), ya

que la fluencia verbal está fuertemente influenciada por el nivel educativo²⁴. Por otra parte, hemos observado que la simple evaluación generaba en las personas entrevistadas niveles significativos de ansiedad que podrían influir en el

resultado de la prueba, al tratarse de un test cronometrado. Algo similar ocurriría con el TR, del que, a pesar de ser un instrumento ampliamente utilizado y con unas correctas sensibilidad y especificidad, diversos autores cuestionan su uso debido a su elevado porcentaje de falsos positivos²⁵ y a sus diferentes sistemas de puntuación.

El TFV y el TR presentan una correlación moderada pero estadísticamente significativa respecto al MEC, resultado que está en consonancia con los de otros autores^{26,27}.

No obstante, la concordancia del TFV y del TR con el MEC es pobre, con índices κ del 33,5 y el 40%, respectivamente ($p < 0,001$), resultado en nada sorprendente si tenemos presente la moderada correlación que presentan. Esta baja concordancia quizás pueda encontrar explicación en el hecho de que los diferentes instrumentos evalúan distintos aspectos de la esfera cognitiva. Mientras el MEC es un test de cribado completo que evalúa varios aspectos de la esfera cognitiva, como orientación, fijación, memoria, praxias, cálculo, lenguaje y abstracción, el TFV se centra más en aspectos frontales (funcionamiento ejecutivo) y mnésicos²⁸, y el TR se centra en los aspectos visuoespaciales y visuoconstructivos, así como también en el funcionamiento ejecutivo²⁹. A este respecto, sería interesante el desarrollo de futuros estudios que analicen esta cuestión.

Respecto al estudio de la validez de los diferentes test de cribado, si se utiliza el MEC como patrón oro, la sensibilidad y especificidad obtenidas en nuestro trabajo son algo menores que las ofrecidas en la literatura médica, si bien los estudios revisados presentan resultados heterogéneos al respecto en dependencia de la muestra y de los puntos de corte utilizados, y la mayoría proceden de población seleccionada. Un resultado común a los 3 instrumentos analizados y especialmente trascendente es que todos ellos parecen clasificar adecuadamente a aquellas personas que dan un resultado negativo, y observan valores predictivos negativos superiores al 90%.

Con respecto a la capacidad diagnóstica de las pruebas analizadas, las áreas bajo la curva ROC obtenidas definen para todas ellas una correcta, aunque moderada, capacidad de discriminación, y estos resultados no se diferencian de trabajos anteriormente publicados, aunque se hayan realizado sobre población seleccionada³⁰.

En relación con las limitaciones de nuestro estudio, es necesario considerar el reducido tamaño de la muestra, especialmente de la utilizada para evaluar el TIN, debido a que esta prueba se realizaba exclusivamente en aquellas personas que acudían acompañadas a la evaluación o que tenían problemas para la realización de ésta. Asimismo, cabe destacar la limitación que supone la elección de un criterio de referencia adecuado que clasifique adecuadamente a los pacientes con DC. La aproximación que se ha tomado en este estudio nos ha parecido la más adecuada teniendo en cuenta los objetivos.

Del análisis de los resultados del trabajo podemos extraer las conclusiones de que todos los test de cribado de DC breves que se han evaluado (TFV, TR y TIN) presentan una validez interna suficiente, que el grado de concordancia entre ellos es moderado, y que el mayor rendimiento diagnóstico, aunque moderado, corresponde al TFV. Para el clínico puede resultar interesante disponer y conocer la rentabilidad diagnóstica de instrumentos breves, sencillos y fáciles de administrar como el TFV, el TR o incluso la

información obtenida de un informador fiable. Aun sin ser una alternativa plenamente válida respecto al MEC, pueden eventualmente ser útiles como test de cribado en las consultas de atención primaria, dadas las circunstancias de limitación temporal en las que se desarrolla la tarea asistencial; aspecto éste que, dadas sus características, incrementa la validez externa de los resultados que presentamos.

Finalmente, los resultados indican algunas reflexiones interesantes para la práctica clínica a la hora de realizar un cribado de DC: por un lado, la necesidad de evaluar habitualmente el nivel educativo de los pacientes, dado que influye decisivamente en el resultado de los test, y se debe, incluso, ajustar la puntuación del TFV en función de ese nivel; por otro lado, la conveniencia de requerir información de un allegado fiable, cuestiones ambas que pueden contribuir al incremento de la validez de los test.

Lo conocido sobre el tema

- El deterioro cognitivo (DC) tiene una elevada prevalencia en personas de más de 64 años, que puede llegar al 20%.
- Los test de cribado de DC son útiles para su aplicación desde la atención primaria.
- Existe gran variabilidad en los resultados de los distintos test de cribado cognitivo.

Qué aporta este estudio

- Los test evaluados presentan una moderada-baja concordancia entre ellos, presumiblemente por explorar distintos aspectos de la esfera cognitiva.
- Se constata el alto valor predictivo negativo de los test de cribado de DC.
- Se observa la conveniencia de evaluar el nivel educativo de los pacientes al realizar test de cribado cognitivos, y la recomendación de complementar el estudio con datos procedentes de informadores-cuidadores fiables.

Conflictos de intereses

Los autores declaran no tener ningún conflicto de intereses.

Financiación

El FIS PI042546 del Instituto de Salud Carlos III (Ministerio de Sanidad y Consumo) ha financiado este estudio.

Bibliografía

1. Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, et al. Mild cognitive impairment. Lancet. 2006;367:1262–70.
2. Lobo A, Saz P, Marcos G, Díaz JL, De la Cámara C. The prevalence of dementia and depression in the elderly community in a Southern European population: The Zaragoza study. Arch Gen Psychiatry. 1995;52:497–506.

3. De Simons LA, McCallum J, Friedlander Y, Simons J. Healthy ageing is associated with reduced and delayed disability. *Age Ageing*. 2000;29:143–8.
4. Ganguli M, Dodge HH, Shen C, DeKosky ST. Mild cognitive impairment, amnestic type: An epidemiologic study. *Neurology*. 2004;63:115–21.
5. Olivera J, Benabarre S, Lorente T, Rodríguez M, Pelegrín C, Calvo JM, et al. Prevalence of psychiatric symptoms and mental disorders detected in primary care in an elderly Spanish population. The Psicotard Study. Preliminary findings. *Int J Geriatr Psychiatry*. 2008;23:915–21.
6. Grupo de Trabajo de Demencias de la semFYC. Demencias desde la Atención Primaria. Barcelona: SemFYC ediciones; 2005.
7. Olivera J, Rodríguez M, Lorente T, Benabarre S, Pardo MA, Ceresuela A, grupo Psicotard. Detección de factores relacionados con el desarrollo de trastornos mentales en población geriátrica (PSICOTARD). *Aten Primaria*. 2006;38:353–7.
8. González MJ, Escrivá R, Vinyoles E, Espel C, Davins J, Borrell M. Estimaciones de la frecuencia de déficit cognitivo según el test empleado. *Aten Primaria*. 1997;20:173–9.
9. Jorm AF. Methods of screening for dementia: A meta-analysis of studies comparing an informant questionnaire with a brief cognitive test. *Alz Dis Assoc Disord*. 1997;11:158–62.
10. Onís MC, Aguado C, Dueñas R, Martínez J, Márquez AB, Fimia A. Validez del diagnóstico de depresión en pacientes ancianos, en consultas de atención primaria. *Aten Primaria*. 1998;22:275.
11. Montón C, Pérez-Echevarría MJ, Campos R, García-Campayo J, Lobo A, el GZEMPP. Escalas de ansiedad y depresión de Goldberg. Una guía de entrevista eficaz para la detección del malestar psíquico. *Aten Primaria*. 1993;12:345–9.
12. Saz P, Díaz JL, De la Cámara C, Carreras S, Marcos G, Lobo A. Reliability and validity of the Spanish version of the GMS-AGECAT package for the assessment of dementia and cognitive disturbances. *Int J Geriatric Psychiatry*. 1996;11:721–8.
13. Linn BS, Linn MW, Gurel L. Cumulative illness rating scale. *J Am Geriatr Soc*. 1968;16:622–6.
14. Bulbena A, Jauregui JV, Zabala MJ. Índice acumulativo de enfermedad. Adaptación castellana del Cumulative illness rating scale en población psicogeriatrica. *Rev Gerontol*. 1996;6:313–8.
15. Lobo A, Saz P, Marcos G, Díaz JL, De la Cámara C, Ventura T, et al. Revalidación y normalización del Mini-Mental Cognoscitivo (primera versión en castellano del Mini-Mental Status Examination) en la población general geriátrica. *Med Clin (Barc)*. 1999;112:767–74.
16. Cacho J, García-García R, Arcaya J, Gay J, Guerrero-Peral AL, Gómez-Sánchez JC, et al. El test del reloj en ancianos sanos. *Rev Neurol*. 1996;24:1525–8.
17. Morales JM, González-Montalvo JL, Bermejo F, Del Ser T. The screening of mild dementia with shortened Spanish version of the “Informant Questionnaire on Cognitive Decline in the Elderly”. *Alz Dis Ass Disord*. 1995;9:105–11.
18. Pascual Millán LF, Martínez Quiñones JV, Modrego Pardo P, Mostacero Miguel E, López del Val J, Morales Asín F. El set-test en el diagnóstico de la demencia. *Neurología*. 1990;5:82–6.
19. Efron B, Gong G. A leisurely look at the bootstrap, the Jackknife and cross-validation. *Am Stat*. 1983;37:36–48.
20. Abraira V, Perez de Vargas A. Generalization of the kappa coefficient for ordinal categorical data, multiple observers and incomplete designs. *Questio*. 1999;23:561–71.
21. 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.
22. Morales JM, González-Montalvo JL, Del Ser T, Bermejo F. Validation of the S-IQCQDE: The spanish version of the information questionnaire on cognitive decline in the elderly. *Arch Neurobiol*. 1992;55:262–6.
23. Schulman KI, Herrman N, Brodaty H, Chiu H, Lawlor B, Ritchie K, et al. IPA survey of brief cognitive screening instruments. *Int Psychogeriatr*. 2006;18:281–94.
24. Da Silva CG, Petersson KM, Faísca L, Ingvar M, Reis A. The effects of literacy and education on the quantitative and qualitative aspects of semantic verbal fluency. *J Clin Exp Neuropsychol*. 2004;26:266–77.
25. Seigerschmidt E, Mösch E, Siemen M, Förstl H, Bickel H. The clock drawing test and questionable dementia: Reliability and validity. *Int J Geriatr Psychiatry*. 2002;17:1048–54.
26. Shulman KI. Clock drawing: Is it the ideal cognitive screening test? *Int J Geriatr Psychiatry*. 2000;15:548–61.
27. Weiner MF, Neubecker KE, Bret ME, Hynan LS. Language in Alzheimer's disease. *J Clin Psychiatry*. 2008;69:1223–7.
28. Parks RW, Loewenstein DA, Dodrill KL, Barker WW, Yoshii F, Chang JY, et al. Cerebral metabolic effects of a verbal fluency test: A PET-scan study. *J Clin Exp Neuropsychol*. 1988;10:565–75.
29. Peters R, Pinto E. Predictive value of the clock drawing test. *Dement Geriatr Cogn Disord*. 2008;26:351–5.
30. Barrero FJ, Vives F, Morales B. Evaluación de la versión española del Memory Impairment Screen. *Rev Neurol*. 2006;43:15–9.