

J. Peña-Casanova^{1,2}
A. Monllau^{1,2}
N. Gramunt Fombuena^{1,2}

La psicometría de las demencias a debate

¹ Sección de Neurología de la Conducta y Demencias
Servicio de Neurología
Hospital del Mar
Instituto Municipal de Asistencia Sanitaria
Barcelona

² Unidad de Investigación en Servicios Sanitarios
Instituto Municipal de Investigación Médica
Barcelona

El diagnóstico preciso de las demencias, así como el cribado de los trastornos cognitivos, es una clave para atender las necesidades de la realidad sociodemográfica. En este artículo se plantean los objetivos del cribado y del diagnóstico, los instrumentos que se utilizan y sus características psicométricas esenciales, especialmente la fiabilidad y la validez. Se destaca la necesidad de evaluar críticamente las publicaciones sobre el tema y se refieren los *Standards for Reporting of Diagnostic Accuracy* (STARD). Además se citan y revisan propuestas y recomendaciones de la literatura. Por último se comentan los tests más estudiados y usados por neurólogos españoles. Destacan el MMSE (*Mini-Mental State Examination*), el MIS (*Memory Impairment Screen*) y el Eurotest.

Palabras clave:
Psicometría. Demencia. Tests.

Neurología 2007;22(5):301-311

Psychometry of dementias at debate

The accurate diagnosis of dementias and the screening of cognitive impairment constitute a key to attend to the sociodemographical needs. This paper states the aims of screening and diagnosis, the tools used and their main psychometric features, especially, reliability and validity. The need to critically review the publications on this subject is highlighted and the Standards for Reporting of Diagnostic Accuracy (STARD) are mentioned. In addition, some proposals and recommendations from literature are cited and reviewed. Finally, the tests studied and used most by Spanish neurologists are commented

on. The MMSE (Mini-Mental Status Examination), the MIS (Memory Impairment Screen) and the Eurotest stand out.

Key words:
Psychometry. Dementias. Tests.

INTRODUCCIÓN

La demencia se define como un síndrome adquirido caracterizado por un déficit cognitivo múltiple suficientemente importante como para interferir en la actividad cotidiana del individuo¹. El diagnóstico se realiza habitualmente en dos pasos: en primer lugar se efectúa el reconocimiento del síndrome demencia y en segundo lugar se realiza el diagnóstico nosológico²⁻⁵. La serie de enfermedades que cursan con demencia es realmente amplia (v. la clasificación de la Sociedad Española de Neurología)⁶, aunque la realidad epidemiológica conduce a los cuadros más frecuentes: enfermedad de Alzheimer, demencia frontotemporal, demencia con cuerpos de Lewy y demencia vascular⁷.

Una buena historia clínica constituye el elemento más importante en el diagnóstico de las demencias, ya que permite reconocer los trastornos del sujeto, la naturaleza de los síntomas y su progresión^{5,8}. Por consiguiente, la historia debe ser cuidadosa, detallada y metódica. En este punto es fundamental el papel de un informador fiable. La historia permite reconocer el binomio del deterioro: por un lado, de las capacidades cognitivas, y por otro, de las actividades funcionales desarrolladas en el contexto de la vida diaria^{3,5}. Esta información se integra con el resto de los datos del paciente: antecedentes, exploración neurológica y general, pruebas complementarias (analítica, neuroimagen, etc.), así como un examen del estado mental que permita constatar y graduar el deterioro de las capacidades cognitivas del individuo⁴. A partir de la historia clínica y de datos de la entrevista, obviamente compatibles con el diagnóstico de demencia, se puede recurrir a categorías descriptivas y situar a los pacientes en distintos grados de posible deterioro. Este es el caso de las categorías de las *Global Deterioration Sca-*

Correspondencia:
Jordi Peña-Casanova
Sección de Neurología de la Conducta y Demencias
Hospital del Mar
Instituto Municipal de Asistencia Sanitaria
Passeig Marítim, 25-29
08003 Barcelona
Correo electrónico: jpcasanova@imas.imim.es

Recibido el 1-6-06
Aceptado el 7-6-06

les⁹ o de la escala CDR (*Clinical Dementia Rating*)¹⁰. Pero, con independencia de la posibilidad de usar categorías generales, la medición del deterioro se realiza, como se ha comentado anteriormente, en los ámbitos cognitivo y funcional (vida diaria)¹¹.

Las pruebas diagnósticas tienen como objetivo deducir el problema de la existencia o no de una enfermedad determinada. En el caso de que se sepa de antemano que el paciente está afecto o no de la enfermedad estudiada, el resultado de la prueba complementaria en cuestión no es relevante para el proceso de diagnóstico. En este punto surge la siguiente pregunta: ¿Cómo realizar adecuadamente la evaluación del trastorno cognitivo?

OBJETIVOS

El presente artículo pretende discutir los estándares metodológicos de la psicometría de las demencias en el contexto del conjunto de escenarios y de datos clínicos que conducen a su diagnóstico. Como se reconoce en un libro reciente sobre tests neuropsicológicos, se trata de incidir en una de las grandes preguntas relacionadas con la metodología: ¿Conoce usted bien sus herramientas? (*How well do you know your tools?*)¹². Además es también crucial responder a lo siguiente: ¿Conoce usted bien los estándares metodológicos de su funcionamiento? Para responder a estas preguntas el presente trabajo seguirá el esquema de un artículo de opinión, en forma de preguntas y respuestas. Por este motivo no se realizará una revisión sistemática del tema en cuestión.

DESARROLLO

Para organizar un debate sobre el tema de la psicometría en el estudio de las demencias conviene empezar definiendo los objetivos de la misma para pasar, posteriormente, al tema de las características de los instrumentos de medición en el contexto clínico. En función de esta caracterización y definición de los estándares metodológicos básicos el lector tendrá, consecuentemente, criterios para analizar si los instrumentos que usa son los más adecuados¹³.

¿Cuáles son los objetivos del diagnóstico? ¿Qué papel tienen las pruebas diagnósticas ante las incertidumbres? El objetivo del diagnóstico es establecer la existencia o la probabilidad de la existencia de una enfermedad en función de un conjunto de signos y síntomas que constituyen un perfil clínico^{14,15}. En el ámbito de la neurología de la conducta, y específicamente de las demencias, el aspecto central es el trastorno cognitivo. Consecuentemente, las pruebas diagnósticas en este ámbito se realizarán para verificar el trastorno cognitivo «en pacientes bajo sospecha de enfermedad» pretendiendo modificar el grado de incertidumbre planteado por un perfil clínico determinado¹⁵. El contrapunto se puede encontrar en los casos en que el deterioro es

obvio a través de la historia clínica. La prueba no aportará disminución de incertidumbres, pero sí permitirá definir un «grado de deterioro» que servirá de índice para posteriores controles evolutivos¹⁶.

¿Test de cribado (screening) o test de diagnóstico? Los instrumentos de cribado (*screening*) (sean biológicos, psicométricos o físicos) pretenden identificar a los sujetos que pueden presentar una enfermedad¹⁷⁻²⁰. En principio no pretenden establecer un diagnóstico. Estas pruebas se aplican en poblaciones que tienen un alto riesgo de padecer un trastorno determinado. Al contrario, las pruebas diagnósticas se aplican para confirmar un diagnóstico en los casos en que una prueba de *screening* ha sido positiva. Un buen test de *screening* debería ser capaz de detectar la mayoría de casos buscados (alta sensibilidad) y presentar pocos falsos positivos (alta especificidad). Entre el cribado y el diagnóstico precoz de los sujetos asintomáticos existen semejanzas, a la vez que claras y cruciales diferencias con el diagnóstico de los sujetos sintomáticos. Sin embargo, existe una línea de unión entre estos tres objetivos^{17,18}. En este sentido el subcomité de estándares de Calidad de la Academia Americana de Neurología (AAN)⁴ estableció que «los instrumentos de cribado son útiles para la evaluación del grado de alteración cognitiva por parte del clínico (guía), al igual que los tests neuropsicológicos (guía), los tests cognitivos focalizados (opción) y ciertas entrevistas estructuradas practicadas a informadores (opción)» (p. 1133). Un test «focalizado» es una prueba que, a diferencia las pruebas que estudian diferentes funciones cognitivas como la memoria, el lenguaje o la orientación, se centra en el estudio de un solo aspecto, como la evocación de nombres de animales.

En este momento es importante comentar un hecho importante. El estándar de trabajo de la AAN sitúa los tests de cribado en paralelo a los tests neuropsicológicos, pero no específica más. En una situación clínica normal, con medios adecuados, se debería realizar un estudio neuropsicológico detallado como continuación de los tests de *screening*¹³. Consecuentemente, los tests neuropsicológicos (como segundo paso psicométrico) se deberán someter, como mínimo, a las mismas exigencias métricas que los tests de cribado. Según esta afirmación, el debate sobre la psicometría de las demencias se debe situar tanto en los instrumentos de *screening* como en los tests neuropsicológicos o las baterías formales de tests. Cabe recordar, sin embargo, que una precisa valoración psicométrica nunca debe realizarse sin considerar la semiología que se desprende de la ejecución de los sujetos.

¿Cómo debe ser un instrumento psicométrico para aplicarse adecuadamente en clínica? Los instrumentos psicométricos deben cumplir una serie de características para que se puedan aplicar adecuadamente a los pacientes¹⁹. En el caso de las demencias existen requisitos especiales^{13,16,17}. Los instrumentos de *screening* deberían tener una serie de características formales como la brevedad (por las limitaciones de tiempo de la clínica), aceptabilidad por parte de los pacientes, simplicidad, facilidad de administración y puntuación y

bajo coste económico. Por último, los tests deberían ser fácilmente adaptables a distintas culturas en la búsqueda de estudios comparativos (adaptabilidad transcultural)¹⁸. Los tests neuropsicológicos no están sometidos a las limitaciones propias del *screening*, pero deberían tener tres requisitos básicos: idoneidad, factibilidad y utilidad¹⁶. Los instrumentos neuropsicométricos deben cumplir, además, una serie de características como instrumentos de medida. Las características métricas son múltiples, pero en la presente aproximación únicamente se consideran brevemente la fiabilidad y la validez.

La *fiabilidad* de un test describe la confianza que se puede depositar en los resultados obtenidos a través de su aplicación. Un instrumento fiable debe ser preciso. Dicho en palabras más técnicas: la fiabilidad se refiere a la proporción de la variación en los resultados que no es consecuencia de errores de medida. Representaría la proporción de la varianza total atribuible a diferencias verdaderas de los individuos. La disminución del error confiere al instrumento un mejor índice de estabilidad y de replicabilidad¹⁷⁻²⁰. La fiabilidad de un instrumento incluye o denota diversos aspectos: consistencia interna, estabilidad (o replicabilidad) y concordancia interjueces.

La *consistencia interna* evalúa la medida en que todos los elementos que conducen a una puntuación final de un cuestionario o una batería reflejan un mismo constructo. La consistencia interna habitualmente se mide calculando el coeficiente alfa de Cronbach o dividiendo el test en dos mitades y comparando las puntuaciones de cada una de ellas¹⁹. El coeficiente alfa de Cronbach viene a representar la media de todas las correlaciones entre cada ítem y la puntuación total de todos los ítems. En cuanto a su graduación se considera que una alfa superior a 0,8 es excelente, superior a 0,7 es buena e inferior a 0,4 es pobre o mala²⁰. En este sentido, por ejemplo, ya que los ítems que componen la escala Bayer-ADL²¹ miden prácticamente lo mismo, es posible obtener la puntuación final a partir de la media aritmética de las puntuaciones con independencia de que algunos ítems no sean aplicables en un paciente dado.

La *fiabilidad interevaluador* (interjueces) e intraevaluador determina el acuerdo entre los resultados de la prueba cuando la realizan varios evaluadores o cuando un mismo evaluador la aplica en momentos distintos¹⁶⁻²⁰. Si estas fiabilidades son bajas las diferencias no reflejarán verdaderas modificaciones en el estado del paciente. Estas fiabilidades frecuentemente se miden mediante un valor kappa, que indica el grado de acuerdo tras una corrección por factores de azar. Los valores de kappa deberían situarse más allá de 0,7 en la evaluación de la fiabilidad interevaluador y superiores a 0,9 para la fiabilidad intraevaluador²⁰. La fiabilidad también se puede medir mediante coeficientes de correlación, pero estos coeficientes pueden ser equívocos. Como alternativa a este coeficiente se puede calcular el coeficiente de correlación interclase. La fiabilidad test-retest (intraevaluador) pretende reflejar la replicabilidad de los resultados cuando

la prueba se administra a la misma persona en momentos distintos¹⁶. Para estudiar adecuadamente esta fiabilidad se debe evitar readministrar la prueba en un corto plazo (para evitar los efectos del recuerdo), o en un plazo de tiempo excesivamente largo (para evitar el efecto de la posible evolución de la enfermedad).

La *validez* de un instrumento de medición neuropsicológica constituye el grado en que mide lo que debe medir y da lugar a resultados similares a los del estándar oro o criterio¹⁷. La validez tiene múltiples dimensiones, entre las que destacan la validez de contenido, la validez de constructo y la validez de criterio. La *validez de contenido* evalúa si el test o cuestionario abarca el elemento, o los elementos, que pretende medir. En este sentido, un test de *screening* de demencia debería como mínimo evaluar la memoria, mientras que un test diagnóstico debería incluir otros componentes cognitivos. La validez de constructo evalúa la medida en que un test se correlaciona con otros tests que miden el mismo dominio cognitivo (validez convergente) y no se correlaciona con pruebas que evalúan otros dominios cognitivos (validez discriminativa). Habitualmente se usa el estadístico kappa para calcular el acuerdo entre pruebas distintas¹⁷.

La *validez de criterio* aporta información sobre la relación del test con el estándar oro. Dicho en otras palabras, la validez de criterio representa el grado de similitud de los resultados del test con la variable o patrón oro. El criterio debe ser siempre una medida independiente y obtenida de forma distinta del test estudiado. Para que este aspecto se cumpla adecuadamente se deben realizar estudios en los que los investigadores sean «ciegos» en relación con el criterio. Existen dos clases de validez de criterio: la validez predictiva y la validez concurrente. El concepto de validez predictiva se refiere a la capacidad de un test para anticipar los resultados de la aplicación del criterio. La validez predictiva es probablemente el mejor método para determinar la validez de los instrumentos neuropsicológicos¹⁹.

La *validez concurrente* normalmente se mide mediante el cálculo de la sensibilidad, la especificidad y el área bajo la curva ROC (*Receiver Operating Characteristic*). La sensibilidad constituye la proporción de pacientes efectivamente enfermos (según el criterio oro) que presentan positividad en la prueba. La especificidad constituye la proporción de sujetos sin la enfermedad y con un resultado negativo en la prueba. La curva ROC permite establecer las relaciones entre la sensibilidad y la especificidad en función de posibles puntos de corte entre la patología y la normalidad. La curva se establece mediante dos ejes, en el eje de ordenadas se representa el grado de verdaderos positivos (sensibilidad), mientras que en el eje de abscisas se representan los falsos positivos (1-especificidad). La curva se traza en función de los distintos puntos de corte. La prueba que resultara con el mismo número de verdaderos positivos y verdaderos negativos daría lugar a una línea recta de 45° y el área bajo esta «curva» sería de 0,5. El instrumento ideal daría un área bajo la curva de 1 (todos los sujetos estarían bien clasificados).

Aunque el conocimiento de la sensibilidad y especificidad diagnósticas de una prueba es importante, habitualmente, en clínica se plantea la siguiente pregunta: si el resultado de la prueba es positivo, ¿cuál es la probabilidad de que el paciente sufra la enfermedad? (este es el valor predictivo positivo de la prueba: VPP). También se puede plantear la siguiente pregunta: si el resultado es negativo, ¿qué probabilidad tiene el paciente de no padecer la enfermedad? (este es el valor predictivo negativo de la prueba: VPN). Estos índices (y otros) no dependen sólo de la sensibilidad y especificidad de la prueba, sino también de la prevalencia (probabilidad anterior) de la enfermedad considerada en la población de estudio.

La *razón de probabilidad (likelihood ratio)* para una prueba determinada establece la probabilidad de que una prueba sea positiva (anormal) ante la patología en estudio. La probabilidad, por ejemplo, de que el *Mini-Mental Status Examination* (MMSE) sea anormal en un paciente con la patología en estudio comparado con la probabilidad que el mismo resultado se observe en un paciente sin dicha patología, se establece a partir de los datos conocidos de su sensibilidad y especificidad en una población determinada en la que se conoce la prevalencia de la enfermedad. Si la prevalencia de la enfermedad es baja (la enfermedad es poco frecuente en la población estudiada), el valor predictivo positivo de una prueba positiva es bajo. En este caso el valor predictivo negativo de una prueba negativa aumenta. Si la prevalencia de la enfermedad es alta (la enfermedad es relativamente frecuente en la población estudiada) el valor predictivo positivo aumenta, mientras que el valor predictivo negativo se reduce. Por este motivo es muy importante conocer los distintos valores predictivos en función de distintas prevalencias^{21,22}.

La *probabilidad postest* establece la probabilidad de que el paciente del que se sospecha que está afecto de una enfermedad determinada y, presentando un valor positivo (anormal) en una prueba dada, sufra efectivamente dicha enfermedad. La oportunidad relativa postest (*postest odds*) se calcula multiplicando la probabilidad pretest (*pretest odds*) por la razón de probabilidad (*likelihood ratio*) de una prueba positiva.

En el contexto del diagnóstico de la enfermedad de Alzheimer es fundamental la consideración de las características cognitivas del envejecimiento. La prevalencia de la EA presenta una fuerte relación positiva con la edad (a más edad, mayor probabilidad de padecer una enfermedad de Alzheimer preclínica). Consecuentemente es lógico que en el estudio del valor diagnóstico de los tests se deban tener en cuenta los factores de probabilidad pretest en función del grupo de edad considerado. En otras palabras: no es lo mismo una puntuación por debajo de un punto de corte determinado en una persona de 65 años que en una persona de 85. *A priori*, la persona de 85 años de edad tiene una mayor probabilidad de padecer una enfermedad demenciante. Por estos motivos el valor predictivo de una prueba se debe estudiar en relación con la prevalencia de la enfermedad en cuestión.

La significación frente al tamaño del efecto. Es importante diferenciar las significaciones estadísticas (habitualmente $p < 0,05$) del tamaño del efecto. El tamaño del efecto (*d*) indica el porcentaje de individuos que se solapan entre el grupo control y el grupo experimental. En términos matemáticos, *d* es la diferencia entre las medias de los pacientes y los controles calibrada en función de unidades de desviación estándar mancomunadas (*pooled*). Un buen marcador diagnóstico debería discriminar aproximadamente a todos los pacientes de los sujetos controles en la variable de interés. Tal discriminación se obtiene si *d* es mayor de 3 (el porcentaje de solapamiento es menor del 5%)²³.

¿Son importantes los datos normativos? Obviamente, los datos normativos son importantes, pero el punto clave para la práctica está, además, en el grado de semejanza entre el paciente en cuestión y el grupo de referencia o normativo disponible. Lo ideal sería disponer de grupos de referencia amplios, con todo tipo de características, de edad, sexo, escolaridad, lengua materna, etc. Cuestiones como el analfabetismo, el bajo nivel educativo y la multiculturalidad añaden particularidades y dificultades a la evaluación cognitiva. La población de la comarca de Nijar en Almería citada por Carnero²⁴ sería un ejemplo de «una sociedad multilingüe y multicultural a la que hay que proporcionar atención sanitaria y a la que habrá que aplicar los instrumentos diagnósticos disponibles» (p. 49). Es tal la importancia de los datos normativos que han aparecido libros que intentan recoger y comentar las publicaciones normativas de los principales tests neuropsicológicos (v., p. ej., Mitrushina et al.²⁵). En este punto es importante recordar que en nuestro medio no existen publicaciones adecuadas sobre normas de tests neuropsicológicos. Por este motivo la «psicometría» que se puede realizar es bastante «intuitiva». En muchos casos no existen normas, o las normas no son adecuadas para el sujeto en estudio. En otras ocasiones, los datos normativos de un test no se corresponden con los datos normativos de otro y, consecuentemente, será difícil o especulativo establecer disociaciones entre funciones. Cabe recordar que en ciertos casos se toman como referencia las tablas normativas de otros países o se citan instrumentos no validados. Así, por ejemplo, en el «Documento Sitges»²⁶ sobre la capacidad de tomar decisiones durante la evolución de una demencia aparece un listado «de algunos de los tests que se utilizan en neurología cognitiva para la valoración de alteraciones intelectuales y psicopatológicas». Un análisis del listado, con independencia de la discutible bibliografía citada, pone de relieve (como refieren los autores) el problema de la falta de datos normativos. No hay alternativas, o se normalizan los tests para nuestra población, o la neuropsicología psicométrica seguirá teniendo componentes claramente especulativos²⁷.

¿Normas comparativas? Las tradicionales normas de los tests (normas comparativas) aportan información sobre medias, puntuaciones estándar y percentiles que permiten situar al paciente en relación con su grupo sociodemográfico considerando, básicamente, la edad, la escolaridad y el sexo.

El conocimiento del efecto de variables como la edad y la escolaridad permite realizar «ajustes» o «correcciones» de las puntuaciones obtenidas. Estas normas no aportan información sobre la validez discriminativa, resultante de la búsqueda de puntos de corte a partir del estudio de sujetos patológicos y de la consideración de factores de prevalencia en la población estudiada. Los puntos de corte se establecen arbitrariamente en función del tipo de distribución de las puntuaciones. La situación ideal sería disponer de datos normativos similares a los del proyecto MOANS (*Mayo Older Americans Normative Study*)²⁸⁻³¹. En este proyecto se realizó una conormalización de una serie de instrumentos neuropsicológicos. La conormalización constituye la normalización simultánea de distintos instrumentos en una misma población y en un mismo proyecto. Esta aproximación es importante para poder establecer disociaciones entre ámbitos cognitivos de forma consistente. La existencia de bancos de datos normativos permite añadir sujetos o reanalizar las estadísticas previas (v. el número monográfico de la revista *The Clinical Neuropsychologist*)³².

¿Existen alternativas ante las normas comparativas? Ante las normas comparativas³³ tradicionales sería conveniente establecer normas diagnósticas. Estas normas pretenden aportar información en la que las puntuaciones de corte y las probabilidades del diagnóstico o clasificación se concatenan con datos sobre la probabilidad pretest (datos epidemiológicos). En otras palabras, se establecen puntos de corte mediante el estudio de la mejor sensibilidad y especificidad y se consideran las posibilidades diagnósticas en función de distintos índices de prevalencia en la población³⁴. La razón de probabilidad (*likelihood ratio*) establece la probabilidad de que una prueba sea positiva (anormal) ante la enfermedad estudiada.

¿Cómo evaluar los artículos sobre diagnóstico y cribado? Esta es, posiblemente, la pregunta más importante. Una vez se han discutido los requisitos básicos se trata ahora de plantear una serie de preguntas propias de la medicina basada en evidencias¹⁷ sobre una prueba diagnóstica concreta. Ciertamente es importante conocer las características métricas^{35,36} de los instrumentos, pero también es importante ver con detalle si los instrumentos cumplen con los requisitos exigidos y si los estudios sobre diagnósticos se han realizado adecuadamente³⁷.

Para determinar si las características de los instrumentos son adecuadas y si existen las pruebas (evidencias) sobre su precisión diagnóstica se debería contestar adecuadamente a cada una de las preguntas de la tabla 1.

En primer lugar se debería comprobar si se han cumplido las reglas que se desprenden de la primera pregunta y subpreguntas (¿son válidas las pruebas sobre la precisión diagnóstica de este instrumento? o, en otras palabras, ¿se ha desarrollado una investigación adecuada?). A continuación se trata de responder a la segunda pregunta (¿tiene este instrumento una capacidad significativa de distinguir en-

Tabla 1	Preguntas fundamentales sobre las pruebas diagnósticas
<p>¿Son válidas las pruebas (evidencias) sobre la precisión diagnóstica de este instrumento?</p> <ul style="list-style-type: none"> ¿Se usó un grupo independiente y ciego de comparación con un estándar oro de diagnóstico? ¿El test diagnóstico se evaluó en un espectro apropiado de pacientes (similar al que se encontrará en la práctica)? ¿Se aplicó el estándar de referencia con independencia del resultado del test? ¿Se ha validado la prueba (grupo de pruebas) en un segundo grupo independiente de pacientes? <p>¿Tiene este instrumento (válido) una capacidad significativa de distinguir entre sujetos que padecen o no este trastorno específico?</p> <p>¿Se puede aplicar este instrumento (test) a un paciente específico?</p> <ul style="list-style-type: none"> ¿La prueba está disponible, es adecuada y precisa en el medio concreto en el que se va a aplicar? ¿Se puede disponer de una estimación clínicamente razonable de la probabilidad pretest del paciente? <ul style="list-style-type: none"> A partir de la experiencia personal, estadísticas de prevalencia, bases de datos de práctica ¿Los pacientes del estudio son similares al propio? ¿Es poco probable que las posibilidades y probabilidades de la enfermedad hayan cambiado desde que se obtuvieron las pruebas? ¿Las probabilidades postest afectarán al manejo del paciente y vamos a ayudarle? <ul style="list-style-type: none"> ¿Van a inducir a iniciar un tratamiento? ¿Las consecuencias de la prueba ayudarán a que el paciente alcance sus objetivos? 	
<p>Basado en Sakett et al. Evidence-based Medicine. How to practice and teach EBM, 2.ª ed. Edinburgh: Churchill-Livingstone, 2000.</p>	

tre sujetos que padecen o no este trastorno específico?). De hecho, todos estos planteamientos han sido formalizados en lo que sería «la arquitectura de la investigación diagnóstica»³⁷ y en las normativas *Standards for Reporting of Diagnostic Accuracy* (STARD)^{38,39} reflejadas en una lista de comprobación con 25 puntos (tabla 2). Mediante la aplicación de estos puntos se avanzará significativamente en la calidad de los trabajos y se reflejará en la práctica clínica^{39,40}.

¿Qué se encuentra en la literatura? ¿Los estudios son correctos? Los trabajos sobre instrumentos diagnósticos son muy frecuentes en la literatura, pero, desgraciadamente, la mayoría de los trabajos se caracterizan por serios problemas metodológicos^{27,36}. Es muy frecuente encontrar variaciones del mismo instrumento (es decir, nada nuevo), variaciones del método de puntuación, estudios con pocos sujetos, que

Tabla 2 Lista de comprobación de STARD para la publicación de estudios sobre precisión diagnóstica

Sección y tema	Ítem	Descripción
Título, resumen y palabras clave	1	Definir el artículo como un estudio de precisión diagnóstica (epígrafe MeSH recomendado «sensibilidad» y «especificidad»)
Introducción	2	Establecer la cuestión u objetivos de estudio, tales como estimar la precisión diagnóstica o comparar la precisión entre pruebas o entre grupos de sujetos (participantes)
Métodos		
Participantes	3	Describir la población de estudio: criterios de inclusión y exclusión, así como los contextos y lugares donde se recogieron los datos
	4	Describir el reclutamiento de los sujetos: ¿se basó en síntomas de inicio, resultados de pruebas previas, o en que a los sujetos se les administrara la prueba índice o el estándar de referencia?
	5	Describir el muestreo de sujetos: ¿se trata de una serie consecutiva de sujetos definida por los criterios de selección de los ítem 3 y 4? Si no es el caso, especificar cómo se seleccionaron
	6	Describir la recogida de datos: ¿Se planificó antes de la realización de las pruebas índice y estándares de referencia (estudio prospectivo) o se realizó posteriormente (estudio retrospectivo)?
Métodos de la prueba	7	Describir el estándar de referencia y su fundamento
	8	Describir las especificaciones técnicas del material y métodos implicados, incluyendo cómo y cuándo se tomaron las medidas, o citar las referencias de las pruebas índice o el estándar de referencia o ambas
	9	Describir la definición y/o la justificación de las unidades de medida, puntos de corte o categorías de los resultados de las pruebas índice y del estándar de referencia
	10	Describir el número, formación y experiencia de las personas que ejecutan y evalúan las pruebas índice y los estándares de referencia
	11	¿Los evaluadores de las pruebas índice y el estándar de referencia eran desconocedores (ciegos) de los resultados de la otra prueba? Describir cualquier otra información clínica disponible para los evaluadores
Métodos estadísticos	12	Describir los métodos de cálculo o comparación de medidas de precisión diagnóstica y los métodos estadísticos utilizados para medir la incertidumbre (p. ej., intervalos de confianza del 95%)
	13	Describir los métodos para calcular la replicación de los resultados, en caso de haberse realizado
Resultados		
Sujetos	14	Referir cuándo se realizó el estudio, incluyendo las fechas de inicio y finalización del reclutamiento
	15	Describir las características clínicas y demográficas de los sujetos (p. ej., edad, sexo, espectro de los síntomas de inicio, comorbilidad, tratamientos actuales y centro de reclutamiento)
	16	Referir el número de sujetos que, cumpliendo los criterios de inclusión, se sometieron o no a las pruebas índice o al estándar de referencia, o a ambas. Describir por qué a los sujetos no se les administró alguna de las pruebas (se recomienda encarecidamente un diagrama de flujo)
Resultados de la prueba	17	Referir el intervalo entre las pruebas índice y el estándar de referencia, así como cualquier tratamiento instaurado entre las mismas
	18	Referir la distribución de la gravedad de la enfermedad (definir los criterios) en los sujetos objeto de estudio, así como la distribución de otros diagnósticos en sujetos que no padecen la enfermedad
	19	Presentar una tabla 2 × 2 de los resultados de la prueba índice incluyendo resultados indeterminados y perdidos. Para variables continuas referir la distribución de los resultados de la prueba en relación con los resultados del estándar de referencia
	20	Referir cualquier efecto adverso al realizar las pruebas índice o el estándar de referencia
Estimaciones	21	Presentar las estimaciones de precisión diagnóstica y las medidas de incertidumbre estadística (p. ej., intervalos de confianza del 95%)
	22	Presentar cómo se trataron los resultados indeterminados, los casos perdidos y los valores extremos (<i>outliers</i>) de las pruebas índice
	23	Referir las estimaciones sobre la variabilidad en la precisión diagnóstica entre evaluadores, centros, o subgrupos de sujetos, si se ha realizado
	24	Referir las estimaciones de la replicación de la prueba, si se han realizado
Discusión	25	Discutir la aplicación clínica de los hallazgos del estudio

La presente lista de comprobación se puede obtener, sin ninguna restricción, en la dirección de internet <http://www.consort-statement.org/stardstatement.htm> checklist (consultado el 11 de junio de 2006). Bossuyt PM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *British Medical Journal* 2003;326:41-4. STARD: *Standards for Reporting of Diagnostic Accuracy*.

no se den probabilidades en función de tasas de prevalencia, ni se calcule el tamaño del efecto³⁶. A pesar de las limitaciones de muchos trabajos, en el año 2001 la AAN recomendaba una serie de instrumentos con altas evidencias (pruebas) de su funcionamiento⁴ y aportaba la bibliografía correspondiente: MMSE⁴¹⁻⁴⁴, *Memory Impairment Screen (MIS)*²², Test Breve del Estado mental (*Short test of Mental Status*)⁴⁵, Test de los 7 Minutos⁴⁶, Test del Dibujo del Reloj⁴⁷⁻⁴⁹ y el Test del Tiempo y el Cambio de Dinero (*Time and Change Test*)⁵⁰. Las revisiones posteriores descartan muchos instrumentos por la falta de datos psicométricos o por muestras inadecuadas^{51,52}.

Algunos de los tests recomendados por la AAN han sido validados en España con mayor o menor refinamiento metodológico: el MMSE, el MIS, el Test del Reloj y el Test del Dinero (que sería una adaptación del Test del Tiempo y el Cambio)⁵². A continuación se presenta una selección no exhaustiva de trabajos españoles.

En la literatura española destacan numerosos trabajos sobre el MMSE⁵³⁻⁵⁸, y en especial las aportaciones sobre ajustes en función de la edad y la escolaridad. El trabajo de Manubens et al.⁵⁵ es especialmente esclarecedor, ya que aportan tablas de percentiles en función de grados de escolarización para sujetos de más de 70 años de edad. La administración de este test requiere entre 7 y 10 min.

El MIS^{59,60} es un test excelente recomendado por la AAN. Ha sido recientemente adaptado al español con la autorización del autor⁵⁹. El estudio de Böhm et al.⁵⁹ incluyó a 403 sujetos divididos en 188 pacientes con demencia y 215 controles. Las curvas ROC mostraron grados excelentes de discriminación de la demencia (0,944), y aún mejores para la enfermedad de Alzheimer (0,978). El punto de corte establecido fue el mismo que en el estudio original (≤ 4). Cabe comentar la alta tasa de prevalencia de demencia en la población estudiada, pero las tablas de probabilidad pretest dan una clara visión del valor diagnóstico en distintas prevalencias. Los datos españoles publicados proceden de unidades especializadas^{59,60}, pero dadas las características simples del test y la semejanza de los datos con el original se puede usar sin problemas. La administración de este test requiere una media aproximada de 4 min.

El Test del Reloj ha sido sometido a distintos estudios españoles⁶¹⁻⁶³ y forma parte del Test de los 7 Minutos. Ninguno de los trabajos publicados (aunque correctos en función de sus objetivos) aporta datos normativos. La administración de este test requiere una media aproximada de 1 min, o menos.

En cuanto al Test de los 7 Minutos destaca el trabajo de Del Ser et al.⁶⁴, que incluye una muestra de 416 ancianos mayores de 70 años participantes en el estudio longitudinal «Envejecer en Leganés». El trabajo es totalmente correcto en relación con los objetivos de los autores: aportar datos nor-

mativos en sujetos a partir de una edad determinada. La administración de este test requiere una media aproximada de 8 min.

El Test de las Monedas⁶⁵, que sigue, en parte, el concepto de la *Making Change Task*, se ha continuado con el Eurotest. El Eurotest ha sido sometido a un estudio preliminar⁶⁶, a un excelente estudio posterior multicéntrico⁶⁷ en el que se ha mostrado su capacidad diagnóstica y a un estudio de fiabilidad⁶⁸. Los rendimientos en el test no están influidos por factores sociodemográficos o educativos. Tiene una buena capacidad diagnóstica, con adecuada validez de contenido y constructo. La administración de este test requiere una media aproximada de algo más de 8 min. El Eurotest, al igual que el MIS original, constituye el modelo metodológico a seguir en el desarrollo de un test.

La evocación categorial (fluencia verbal de palabras de una categoría determinada o en forma del Set-test) también se ha estudiado en el ámbito normativo, predictivo y de diagnóstico de demencia⁶⁹⁻⁷³. Además dicha prueba está incluida en el Test de los 7 Minutos. Los estudios del tamaño del efecto no muestran excelentes resultados²³.

El Test de Pfeiffer (*Short Portable Mental Status Questionnaire*, SPMSQ) como alternativa al MMSE ha sido validado en España⁷⁴. La administración de este test requiere unos 2 a 3 min. El estudio español reciente muestra una sensibilidad del 85% y una especificidad del 79%.

En este contexto de problemas, el proyecto NORMACODEM⁷⁵ planteó adaptar, normalizar y validar «instrumentos interrelacionados» que evaluaran el deterioro neuropsicológico, su intensidad, su estructura cognitiva básica y la correlacionaran con grados de discapacidad funcional. El proyecto NORMACODEM⁷⁵ incluyó a 254 sujetos control, 86 casos de deterioro sin demencia y 111 casos de enfermedad de Alzheimer. En el momento actual se han publicado trabajos sobre los siguientes instrumentos: *Alzheimer's Disease Assessment Scale (ADAS)*⁷⁶, *Interview for Deterioration of Daily living in Dementia (IDDD)*⁷⁷, MMSE⁵⁷, Test Barcelona⁷⁸, Escala de Blessed (*Blessed Dementia Rating Scale*)⁷⁹, *Geriatric Evaluation of Relative's Rating Instrument (GERRI)*⁸⁰, *Rapid Disability Rating Scale-2 (RDRS-2)*⁸¹. La posibilidad de combinar datos de tests cognitivos, escalas funcionales y de los perfiles del Test Barcelona dan una especial dimensión a este proyecto. No se han publicado datos diagnósticos de la combinación de diversas pruebas.

Por su parte, la ADAS ha sido estudiada por varios grupos^{76,82,83}, e incluso validada en un proyecto de armonización europea de instrumentos para la evaluación de la demencia, el proyecto Euro-Harpid⁸⁴. El *Cambridge Examination for Mental Disorders of the Elderly (CAMDEX)* también está disponible en español⁸⁵⁻⁸⁷ y también está disponible en una versión con validación europea (proyecto Euro-Harpid)⁸⁸.

También existen importantes datos multicéntricos españoles sobre la escala Bayer-ADL en el contexto de un

estudio realizado en tres países europeos²¹. La muestra española, de 619 casos, mostró que tenía una mayor capacidad diagnóstica que el MMSE. Se cita este trabajo porque pone de manifiesto que una buena escala de actividades avanzadas de la vida diaria puede tener un mayor valor predictivo que un test cognitivo.

Por último, cabe citar la serie de tests que, con mayor o menor corrección metodológica, aparece en la literatura internacional. El lector encontrará su revisión en un trabajo de Lorenz, Scanlan y Borson (2002)⁵¹. El *Syndrom Kurztest (SKT)*^{89,90} se incluye en la citada revisión. A este conjunto de tests cabría añadir el MoCA (*Montreal Cognitive Assessment*)⁹¹, un test muy reciente que, aunque existe en versión castellana, no ha sido validado en controles y pacientes españoles.

¿Qué instrumentos psicométricos se pueden aconsejar en nuestro medio? Aunque en el mundo se han publicado múltiples tests, ninguno ha sido reconocido como estándar internacional⁴. De forma ideal se debería aconsejar el test que cumpliera con todos los requisitos enunciados previamente y que estuviera normalizado y validado en la población a la que pertenece el paciente. Una vez establecido este principio se debe descender al pragmatismo y aconsejar el test que más se aproxime al ideal. Puede acontecer que un test haya sido adaptado culturalmente a una población pero no haya sido estudiado adecuadamente en cuanto a su validez. Cabe hacer una serie de comentarios sobre los tests más estudiados y usados por los neurólogos españoles:

- *MMSE*. Probablemente sería el test de elección si se dispone del tiempo suficiente, ya que aporta datos en distintos ámbitos cognitivos. Es adecuado si el paciente tiene un mínimo de escolaridad y/o se aplican ajustes por edad y escolaridad. Convendría tomar como referencia el estudio normativo más adaptado a la realidad del paciente estudiado. Como alternativa más breve se puede usar el SPMSQ de Pfeiffer, siguiendo los datos de la validación de Martínez de la Iglesia et al.⁷⁴. El problema está en el tiempo necesario para su administración.
- *MIS*. Test excelente si se dispone de muy poco tiempo. Durante el tiempo de interferencia se puede realizar, por ejemplo, una prueba de evocación categorial. Una variación del MIS, con imágenes, que está en fase de desarrollo, es el «Test de la Fotos»⁹².
- *Test del Reloj*. Poco aconsejable. Se debería usar cuando el sujeto tenga suficiente escolaridad (estudios primarios completos) y experiencia suficiente en tareas de papel y lápiz. El Test del Reloj no es realmente el mejor instrumento con los datos psicométricos disponibles. Tampoco es un buen test si se usa aislado, ya que «solamente» mide praxis constructiva y ejecución.
- *Test de Fluencia Verbal*. Se debería usar como segunda opción. Los tests de fluencia verbal se pueden utilizar si se realizan ajustes por edad y escolaridad. Tie-

ne mucho interés si se realiza junto con otras pruebas como un test de memoria. La suma de fallos de memoria (p. ej., en el MIS) más los fallos en un test de fluencia verbal son orientativos de demencia.

- *Test de los 7 Minutos*. Se debería usar cuando los sujetos a estudiar sean similares al grupo normativo español publicado. No cumple en absoluto los requisitos para su uso en una consulta de medicina general o de neurología (realmente se requiere más de 7 min, y se requiere material). Faltan datos normativos amplios.
- *Eurotest*. Excelente test para demencia si no se quiere incidir específicamente en la memoria y se dispone de 8 min. Test fiable y excelentemente validado.

Una serie de tests ya publicados muestran unas características de contenidos, métricas y de brevedad que les confieren cierto futuro. Estos tests son los siguientes: Mini-Cog⁹³, el GPCOG⁹⁴ y el 6 ítem Screener⁹⁵ (el lector puede consultar las correspondientes referencias y la revisión crítica de Lorentz et al.⁵¹).

El siguiente paso, el de la exploración neuropsicológica a través de la selección de una serie de tests, se enfrenta a un grave problema: la ausencia de datos normativos sistemáticos españoles. Además, idealmente, se debería disponer de instrumentos connormalizados siguiendo el modelo MOANS. Aunque existen datos normativos españoles diversos (v. la reciente revisión de Peña-Casanova et al.)⁵², no existen, por ahora, trabajos sistemáticos de connormalización en la población en la que la demencia es prevalente. La connormalización es importante, ya que si los datos normativos proceden del mismo grupo de sujetos, las disociaciones entre los rendimientos en los distintos tests adquirirán un valor más objetivo. En el ámbito de los tests también sería aconsejable aplicar los conceptos de las normas diagnósticas, es decir, ver su rendimiento a partir de puntos de corte definidos por grupos de pacientes. Tomar como modelo sistemático la enfermedad de Alzheimer es realmente un problema, ya que ciertos tests específicos, como los tests de lenguaje o ejecutivos, se deberían estudiar en pacientes con lesiones cerebrales focales.

En este contexto sería muy importante estudiar los rendimientos de los tests de *screening* y de los tests neuropsicológicos en grupos normativos, sujetos diagnosticados de trastorno cognitivo leve y sujetos con demencia discreta y moderada. La correlación con datos de neuroimagen y genética sería un complemento importante. Por otro lado, los tests se deberían estudiar en otras situaciones clínicas como la depresión, la ansiedad o trastornos psicóticos.

CONCLUSIONES

La medición de las capacidades cognitivas en el ámbito de las demencias no es una tarea crucial. Como cualquier

prueba médica, de alta o de baja tecnología, la psicometría debe cumplir con los requisitos métricos y diagnósticos adecuados. En los últimos años se han establecido unos estándares para el desarrollo de estudios sobre el diagnóstico y métodos para realizar una publicación completa y precisa de los resultados de los estudios. Estas aportaciones se deberían adoptar sistemáticamente y deberían tenerse en cuenta en el desarrollo de nuevos tests o en la reevaluación de los existentes. El desarrollo del Eurotest es un ejemplo de trabajo de calidad. La aplicación de las guías de trabajo y publicación va a redundar en la calidad científica de los estudios y en una mayor aplicabilidad clínica. La selección de un instrumento se debería hacer a partir de la consideración de los instrumentos disponibles y la realidad del paciente (edad, escolaridad, deficiencias sensoriales, etc.). El *know your tools* (conozca sus instrumentos) es fundamental, pero no se debería olvidar la importancia de la coherencia entre la historia clínica, las manifestaciones funcionales y los resultados de los tests, atendiendo, asimismo, a la semiología. Si en el ámbito del cribado (*screening*) quedan muchos temas pendientes, en el ámbito de los tests neuropsicológicos formales el recorrido restante es aún mayor.

BIBLIOGRAFÍA

- American Psychiatric Association. Manual diagnóstico y estadístico de los trastornos mentales, texto revisado (DSM-IV-TR). Barcelona: Masson, 2001.
- McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology* 1984;34:939-44.
- Corey-Bloom J, Thal LJ, Galasko D, Folstein M, Drachman D, Raskind M, et al. Diagnosis and evaluation of dementia. *Neurology* 1995;45:925-35.
- Petersen RC, Stevens JC, Ganguli M, Tangalos EG, Cummings JL, DeKosky ST. Practice parameter: early detection of dementia: mild cognitive impairment (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 2001;56:1133-42.
- Galdos L, Manubens JM. Historia clínica, exploración física y neurológica y pruebas complementarias. En: Grupo de Estudio de Neurología de la Conducta y Demencias. Sociedad Española de Neurología. Guías en Demencias. Barcelona: Masson, 2003; p. 189-94.
- Robles A, Carnero C, Pascual LF, Alom J, Dobato JL, Romero J, et al. Clasificación. En: Grupo de Estudio de Neurología de la Conducta y Demencias. Sociedad Española de Neurología. Guías en Demencias. Barcelona: Masson, 2003; p. 29-38.
- Knopman DS, DeKosky ST, Cummings JL, Chui H, Corey-Bloom J, Relkin N, et al. Practice parameter: diagnosis of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 2001;56:1143-53.
- Chui H. Clinical history and neurological signs. En: Qizilbash N, Schneider LS, Chui H, Tariot P, Brodaty, Kaye J, Erkinjuntti T, editores. Evidence-based dementia practice. Oxford: Blackwell Publishing, 2003; p. 119-24.
- Reisberg B, Ferris SH, de Leon MJ, Crook T. The Global Deterioration Scale for assessment of primary degenerative dementia. *Am J Psychiatry* 1982;139:1136-9.
- Hughes CP, Berg L, Danzinger WL. A new clinical scale for the staging of dementia. *Br J Psychiatr* 1988;140:566-72.
- Bouchard RW, Rossor MN. Typical clinical features. En: Gauthier S, editor. Clinical diagnosis and management of Alzheimer's disease, 2.^a ed. revisada. Londres: Martin Dunitz, 1996.
- Strauss E, Sherman E, Spreen O. A compendium of neuropsychological tests, 3.^a ed. New York: Oxford University Press, 2006.
- Qizilbash N, Schneider LS, Chui H, Tariot P, Brodaty, Kaye J, Erkinjuntti T. Evidence-based dementia practice. Oxford: Blackwell Publishing, 2003.
- Peña-Casanova J. Alzheimer Disease Assessment Scale-Cognitive in clinical practice. *International Psychogeriatr* 1997;9(Suppl. 1): 105-14.
- Gómez A. Caracterización de las pruebas diagnósticas. *Neurología* 2004(Suppl. 1):31-8.
- Lezak MD, Howieson DB, Loring DW. Neuropsychological assessment, 4.^a ed. New York: Oxford University Press, 2004.
- Sackett DL, Strauss SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-based Medicine, Edinburg: Churchill Livingstone, 2000.
- Shulman K, Feinstein A. Quick Cognitive Screening for Clinicians. London: Martin Dunitz, 2003.
- Retzlaff PD, Gibertini M. Neuropsychometric issues and problems. En: Vanderploeg RD, editor. Clinician's guide to neuropsychological assessment. Hillsdale: Lawrence Erlbaum Associates, 1994; p. 185-209.
- Gifford DR, Cummings JL. Evaluating dementia screening tests. *Neurology* 1999;52:224-7.
- Erzigkeit H, Leffeld H, Peña-Casanova J, Bieber F, Yekrangi C, Rupp M, et al. The Bayer-Activities of Daily Living Scale (B-ADL): results from a validation study in three European countries. *Dement Geriatr Cogn Disord* 2001;12:348-58.
- Buschke H, Kuslansky G, Katz M, Stewart WF, Sliwinski MJ, Eckholdt HM, et al. Screening for dementia with the Memory Impairment Screen. *Neurology* 1999;52:231-8.
- Zakzanis KK, Leach L, Kaplan E. Neuropsychological Differential Diagnosis. Lisse: Swets and Zeitlinger, 1999.
- Carnero C. Analfabetismo, bajo nivel educativo y muestras normativas para pruebas neuropsicológicas. *Neurología* 2005;20:49.
- Mitrushina M, Boone KB, Razani J, D'Elia LF. Handbook of normative data for neuropsychological assessment, 2.^a ed. New York: Oxford University Press, 2005.
- Boada M, Robles A. Análisis y reflexiones sobre la capacidad para tomar decisiones durante la evolución de una demencia: Documento Síctges. Barcelona: Glosa, 2005.
- Peña-Casanova J, Gramunt N, Gich J. Test neuropsicológicos. Fundamentos para una neuropsicología basada en evidencias. Barcelona: Masson, 2004.
- Ivnik RJ, Malec JF, Smith GE, Tangalos EG, Petersen RC, Kokmen E, et al. Mayo's Older Americans Normative Studies: WAIS-R norms for ages 56 to 97. *Clin Neuropsychol* 1992;6(Suppl.):1-30.
- Ivnik RJ, Malec JF, Smith GE, Tangalos EG, Petersen RC, Kokmen E, et al. Mayo's Older Americans Normative Studies: WMS-R norms for ages 56 to 94. *Clin Neuropsychol* 1992;6(Suppl.):49-82.

30. Ivnik RJ, Malec JF, Smith GE, Tangalos EG, Petersen RC, Kokmen E, et al. Mayo's Older Americans Normative Studies: Updated AVLT norms for ages 56 to 97. *Clin Neuropsychol* 1992;6(Suppl.): 83-104.
31. Ivnik RJ, Malec JF, Smith GE, Tangalos EG, Petersen RC. Neuropsychological tests' norms above age 55: COWAT, BNT, MAE Token, WRAT-R Reading, AMNART, Stroop, TMT, and JLO. *Clin Neuropsychol* 1996;10:262-78.
32. Steinberg B, Bieliauskas L. Introduction to the special edition: IQ-based MOANS norms for multiple neuropsychological instruments. *Clin Neuropsychol* 2005;19:277-9.
33. Smith GE, Ivnik RJ. Normative neuropsychology. En: Petersen R, editor. *Mild cognitive impairment. Aging to Alzheimer's disease*. New York: Oxford University Press, 2003; p. 63-88.
34. Sliwinski M, Lipton R, Buschke H, Wasylshyn C. Optimizing cognitive tests norms for detection. En: Petersen RC, editor. *Oxford: Oxford University Press, 2003; p. 89-104*.
35. Serra-Mayoral A, Peña-Casanova J. Fiabilidad test-retest e interevaluador del test Barcelona. *Neurología* 2006;21:277-81.
36. Peña-Casanova J, Böhm P, Manero RM. La evaluación neuropsicológica en la enfermedad de Alzheimer en una medicina basada en evidencias: notas preliminares. *Neurología* 2000;15(Suppl. 4): 31-8.
37. Sackett DL, Haynes RB. Evidence base of clinical diagnosis: the architecture of diagnostic research. *Br Med J* 2002;324:539-41.
38. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. for the STARD steering group. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Br Med J* 2003;326:41-4.
39. Burns DE. The STARD initiative and reporting of studies of diagnostic accuracy. *Clin Chem* 2003;49:19-20.
40. Carnero-Pardo C. Evaluación de las pruebas diagnósticas. *Rev Neurol* 2005;40:641-3.
41. Ganguli M, Belle S, Ratcliff G. Sensitivity and specificity for dementia of population-based criteria for cognitive impairment: the movies project. *J Gerontol* 1993;48:152-61.
42. Kukull WA, Larson EB, Teri L. The Mini-Mental State Examination score and the clinical diagnosis of dementia. *J Clin Epidemiol* 1994;47:1061-7.
43. Wilder D, Cross P, Chen MPJ. Operating characteristics of brief screens of dementia in a multicultural population. *Am J Geriatr Psychiatry* 1995;3:96-107.
44. Tangalos EG, Smith GE, Ivnik RJ. The Mini-Mental State Examination in general medical practice: clinical utility and acceptance. *Mayo Clin Proc* 1996;71:829-37.
45. Kokmen E, Smith GE, Petersen RC. The short test of mental status: correlations with standardized psychometric testing. *Arch Neurol* 1991;48:725-8.
46. Solomon PR, Hirschhoff A, Kelly B. A 7-Minute neurocognitive screening battery highly sensitive to Alzheimer's disease. *Arch Neurol* 1998;55:349-55.
47. Cahn D, Salmon D, Monsch A. Screening for dementia of the Alzheimer type in the community: the utility of the Clock Drawing Test. *Arch Clin Neuropsychol* 1996;11:529-39.
48. Bourke J, Castleden CM, Stephen R. A comparison of clock and pentagon drawing in Alzheimer's disease. *Int J Geriatr Psychiatry* 1995;10:703-5.
49. Lam L, Chiu H, Ng K. Clock-face drawing, reading and setting tests in the screening of dementia in Chinese elderly adults. *J Gerontol* 1998;53:353-7.
50. Inouye SK, Robison JT, Froehlich TE. The Time and Change Test: a simple screening test for dementia. *J Gerontol* 1998;53A:281-6.
51. Lorentz WJ, Scanlan JM, Borson S. Brief screening tests for dementia. *Can J Psychiatry* 1947;8:723-33.
52. Peña-Casanova J. Enfermedad de Alzheimer: los tests de cribado. En: Martínez-Lage M, editor. *Curso Nacional Alzheimer (en prensa)*.
53. Lobo A, Saz P, Marcos G, Día JL, de la Cámara C, Ventura, et al. Revalidación y estandarización del MEC (primera versión española del Mini-Mental Status Examination) en la población geriátrica general. *Med Clin* 1999;112:767-41.
54. Bermejo F, Gómez-Isla T, Morales JM. El Mini-Mental State Examination en la evaluación del deterioro cognitivo y la demencia. En: Del Ser T, Peña-Casanova, editores. *Evaluación neuropsicológica y funcional de la demencia*. Barcelona: JR Prous Editores, 1994; p. 93-7.
55. Manubens JM, Martínez-Lage P, Martínez-Lage JM, Larumbe R, Muruzábal J, Martínez-González, et al. Variación de las puntuaciones en el Mini-Mental-State con la edad y el nivel educativo. Datos normalizados en la población mayor de 70 años de Pamplona. *Neurología* 1998;13:111-9.
56. Lobo A, Marcos G, Día JL, De la Cámara C, Ventura T, Morales Asín F. Validación y estandarización del mini-examen cognoscitivo (primera edición española del Mini-Mental Status Examination) en la población geriátrica general. *Med Clin* 1999;112: 767-74.
57. Blesa R, Pujol M, Aguilar M, Santacruz P, Bertrán-Serra I, Hernández G, et al. Clinical validity of the Mini-Mental State for Spanish speaking communities. *Neuropsychologia* 2001;39: 1150-7.
58. Vilalta-Franch J, Llinás-Regla J, López-Pousa S. Resultados del Mini-Examen Cognitivo como un instrumento de muestreo en los estudios epidemiológicos de demencia. *Neurología* 1996;11:166-9.
59. Böhm P, Peña-Casanova J, Gramunt N, Manero RM, Terrón C, Quiñones-Úbeda S. Versión española del Memory Impairment Screen (MIS): datos normativos y validez discriminativa. *Neurología* 2005;20:402-11.
60. Pérez Martínez DA, Baztán JJ, González-Becerra M, Socorro A. Evaluación de la utilidad diagnóstica de una adaptación española del Memory Impairment Screen de Buschke para detectar demencia y deterioro cognitivo. *Rev Neurol* 2005;40:644-8.
61. Navarro JAG, García PG, Mora J, Pastor E, Gil Gregorio P. El test cognitivo del reloj: utilidad para medir deterioro cognitivo y apraxia. *Rev Esp Geriatr Gerontol* 1995;30:27-31.
62. Martínez Arán A, Solé P, Salameró M, de Azpiazu P, Tomás S, Marín R. El test del dibujo del reloj: métodos de evaluación cuantitativos y cualitativos. *Rev Neurol* 1998;27:55-9.
63. Cacho J, 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.
64. Del Ser T, Sánchez F, García MJ, Otero A, Zunzunegui MV, Muñoz DG. Versión española del Test de los 7 minutos. Datos normativos de una muestra poblacional de ancianos de más de 70 años. *Neurología* 2004;19:344-58.
65. Carnero C, Lendinez A, Navarro E. El test de las monedas. *Rev Neurol* 1999;29:801-4.

66. Carnero-Pardo C, Montoro-Ríos MT. Evaluación preliminar de un nuevo test de cribado de demencias (Eurotest). *Rev Neurol* 2004;38:201-9.
67. Carnero-Pardo C, Gurpegui M, Sánchez-Cantalejo E, Frank A, Mola S, Barquero S, et al., and the Trans-EUROTEST Group. *BMC Neurology* 2006, 6:15 <http://www.biomedcentral.com/1471-2377/6/15>.
68. Carnero-Pardo C, Gurpegui M, Sánchez-Cantalejo E, Gómez A, Guardado P, Montoro T, en nombre del grupo Trans-Eurotest. Evaluación de la fiabilidad del Eurotest. *Neurología* 2007;22:153-8.
69. Pascual LF, Martínez JV, Modrego P, Mostacero E, López del Val J, Morales F. El set test en el diagnóstico de demencia. *Neurología* 1990;5:82-5.
70. Pascual LF, Martínez JV, Modrego P, López del Val LJ, Mostacero M, Morales F. Contribución del set-test en el diagnóstico de la enfermedad de Alzheimer. *Neurología* 1991;6:78-83.
71. Carnero C, Maestre J, Marta J, Mola S, Olivares J, Sempere AP. Validación de un modelo de predicción de la fluencia verbal semántica. *Rev Neurol* 2000;30:1012-5.
72. Alberca R, Salas D, Pérez-Gil JA, Lozano P, Gil-Neciga E. Fluencia verbal y enfermedad de Alzheimer. *Neurología* 1999;15:344-8.
73. Benito-Cuadrado MM, Esteba-Castillo S, Böhm P, Cejudo-Bolibar J, Pena-Casanova J. Semantic verbal fluency of animals: a normative and predictive study in a Spanish population. *J Clin Exp Neuropsychol* 2002;24:1117-22.
74. Martínez de la Iglesia J, Dueñas R, Onís M, Aguado C, Albert C. Adaptación y validación al castellano del cuestionario de Pfeiffer (SPMSQ) para detectar la existencia de deterioro cognitivo en personas mayores de 65 años. *Med Clin* 2001;117:129-34.
75. Peña-Casanova J, Aguilar M, Bertrán-Serra I, Santacruz P, Hernández G, Insa R, et al. Normalización de instrumentos cognitivos y funcionales para la evaluación de la demencia (NORMACODEM): objetivos, contenidos y población. *Neurología* 1997;11:61-7.
76. Peña-Casanova J, Aguilar M, Santacruz P, Bertran-Serra I, Hernández G, Sol JM, et al. Adaptación y normalización españolas de la Alzheimer's Disease Assessment Scale (NORMACODEM) (y II). *Neurología* 1997;12:69-77.
77. Böhm P, Peña-Casanova J, Aguilar M, Hernández G, Sol JM, Blesa R. Clinical validity and utility of the interview for deterioration of daily living in dementia for Spanish-speaking communities. *Int Psychogeriatr* 1998;10:261-70.
78. Peña-Casanova J, Monllau A, Böhm P, Blesa R, Aguilar M, Sol JM, et al. Correlación cognitivo-funcional en la de demencia tipo Alzheimer: a propósito del Test Barcelona abreviado. *Neurología* 2005;20:4-8.
79. Peña-Casanova J, Monllau A, Böhm P, Aguilar M, Sol JM, Hernández G, et al., y Grupo NORMACODEM. Valor diagnóstico y fiabilidad test-retest de la Escala de demencia de Blessed (BDRS) para la enfermedad de Alzheimer: datos del proyecto NORMACODEM. *Neurología* 2005;20:349-55.
80. Soler T, Pujol J, Peña-Casanova J, Hernández G, Sol JM, Aguilar A, et al., and the NORMACODEM Group. Adaptation and standardization of the Geriatric Evaluation of Relative's Rating Instrument (GERRI) for Spain. *Neurología* 2002;17:193-9.
81. Monllau A, Aguilar M, Peña-Casanova J, Böhm P, Sol JM, Hernández G y grupo NORMACODEM. Estudio de la Escala de Evaluación de Rápida de la discapacidad-2 (The Rapid Disability Rating Scale-2) en la enfermedad de Alzheimer: datos del proyecto NORMACODEM. *Neurología* 2006;21:282-8.
82. Manzano JM, Llorca G, Ledesma A, López JJ. Adaptación española de la Alzheimer's Disease Assessment Scale (ADAS). *Actas Luso-Esp Neurol Psiquiatr* 1994;22:64-70.
83. Pascual LF, Saz P, Larumbe R, Martínez-Lage P, Muruzabal J, Morales F, et al. Estandarización en una población española de la escala ADAS (Alzheimer's Disease Assessment Scale). *Neurología* 1997;12:238-44.
84. Verhey FR, Houx P, Van Lang N, Huppert F, Stoppe G, Saerens J, et al. Cross-national comparison and validation of the Alzheimer's Disease Assessment Scale: results from the European Harmonization Project for Instruments in Dementia (EURO-HARPID). *Int J Geriatr Psychiatry* 2004;19:41-50.
85. Vilalta J, Llinás J, López-Pousa S, Amiel J, Vidal C. CAMDEX. Validación de la adaptación española. *Neurología* 1990;5:117-20.
86. Llinás J, Vilalta J, López-Pousa S. Camdex: adaptación y validación españolas. Barcelona: Áncora, 1991.
87. Lozano-Gallego M, Vilalta-Franch J, Llinás-Reglà J, López-Pousa S. El Cambridge Cognitive Examination como instrumento de detección de demencia. *Rev Neurol* 1999;28:348-52.
88. Verhey FR, Huppert F, Korten ECC, Houx P, de Vugt M, Van Lang N, et al. Cross-national comparisons of the Cambridge Cognitive Examination-revised: the CAMCOG-R. *Age Ageing* 2003;32:534-40.
89. Erzigkeit H. The development of the SKT project. En: Hindmarch I, Hippus H, Vilcox J, editores. *Dementia: molecules, methods and measures*. Chichester: John Wiley, 1991; p. 101-8.
90. Erzigkeit H. Manual zum SKT. Formen A-E vorläufiges Manual. Vless: Vatersteten-München, 1989.
91. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. Montreal Cognitive Assessment (MoCA): a brief screening tool for mild cognitive impairment. *J Am Geriatr* 2005;53:695-9.
92. Carnero-Pardo C, Montoro-Ríos MT. Test de las fotos. *Rev Neurol* 2004;39:801-6.
93. Borson S, Scanlan JM, Brush M, Vitalino PP, Dokman A. The mini-cog: a cognitive «vital signs» measure for dementia screening in multi-lingual elderly. *Int J Geriatr Psychiatry* 2000;15:1021-7.
94. Brodaty H, Pond D, Kemp NM, Luscombe G, Harding L. The GPCOG: a new screening test for dementia designed for general practice. *J Am Geriatr Soc* 2002;50:530-4.
95. Callahan CM, Unverzagt FW, Hui SL, Perkins AJ, Hendrie HC. Six-item screener to identify cognitive impairment among potential subjects for clinical research. *Med Care* 2002;40:771-81.