



ORIGINAL

¿Cuál es la combinación de síntomas según padres y maestros más fiable para el diagnóstico de TDAH?

E. Cardo^{a,b,*}, M. Bustillo^a, L. Riutort^a, M.M. Bernad^a, V. Meisel^a, G. García-Banda^b y M. Servera^b

^aLaboratorio de Neurociencias, DEVPSY IUNICS, Hospital Son Llátzer, Palma de Mallorca, Baleares, España

^bDEVPSY IUNICS, Departamento de Psicología, Universidad Islas Baleares, Palma de Mallorca, Baleares, España

Recibido el 16 de febrero de 2009; aceptado el 19 de abril de 2009

Disponible en Internet el 3 de junio de 2009

PALABRAS CLAVE

Trastorno por déficit de atención e hiperactividad;
Niño;
Manual Diagnóstico y Estadístico de los Trastornos Mentales-IV;
Kappa;
Valor predictivo de los test;
Sensibilidad;
Especificidad

Resumen

Introducción: El método politético, recogido en el Manual Diagnóstico y Estadístico de los Trastornos Mentales-IV, es el que se plantea tradicionalmente para el diagnóstico de los trastornos por déficit de atención e hiperactividad (TDAH). Sin embargo, es probable que este procedimiento, que conlleva establecer cualquier combinación de 6 ítems, no sea el más adecuado para establecer un diagnóstico de TDAH y que las diferentes combinaciones de ítems tengan desigual peso.

Objetivo: Determinar qué combinaciones de ítems de las escalas de inatención e hiperactividad-impulsividad de DuPaul, para padres y maestros, son las más efectivas a la hora de diagnosticar o descartar un TDAH.

Resultados: No todas las combinaciones de ítems de padres y maestros obtuvieron el mismo valor predictivo. Todas ellas ofrecieron altos grados de especificidad, pero bajos grados de sensibilidad; es decir, las combinaciones resultaron más efectivas y confiables para descartar este tipo de trastornos que a la hora de diagnosticarlos.

Discusión: El método que implica establecer combinaciones de los ítems mejores predictores del TDAH ofreció superiores resultados que los del método politético. No obstante, se requieren futuros estudios con muestras que posean tasas base más elevadas para obtener datos más fiables. Asimismo, sería conveniente emplear un método alternativo para la selección de los sujetos participantes que permita comprobar, posteriormente, la eficacia diagnóstica pura de los diferentes ítems de las escalas de clasificación de los TDAH.

© 2009 Asociación Española de Pediatría. Publicado por Elsevier España, S.L. Todos los derechos reservados.

*Autor para correspondencia.

Correo electrónico: ecardojalon@gmail.com (E. Cardo).

KEYWORDS

Attention Disorders with or without Hyperactivity; Child; Diagnostic and Statistical Manual of Mental Disorders-IV; Kappa; Predictive Value of Tests; Sensitivity; Specificity

What combination of symptoms according to parents and teachers would be more reliable for the diagnosis of ADHD?

Abstract

Introduction: The polythetic method used in the DSM is the one proposed traditionally for the diagnosis of Attention Disorders with or without Hyperactivity (ADHD). However, it is possible that the approach which aggregates any combination of 6 items won't be the optimal method to establish a diagnosis of ADHD, and that the different combinations may not be the same as regards to their ability to predict ADHD.

Aim: Determine which combinations of items of DuPaul's inattention and hyperactivity-impulsivity scales (from parent and teacher versions), are the most effective to predict or rule out a diagnosis of ADHD.

Results: Not every combination of items from parents and teachers obtained the same predictive value. All of them offered high levels of specificity, but had low sensitivity; that is to say, the combinations were more effective and reliable for ruling out the disorder than predicting it.

Conclusions: Data show that not every combination of ADHD items has the same predictive value and, therefore, the well-known polythetic method is disputable. The highest predictive value combinations, limitations of the study, and future lines of investigation are analyzed.

© 2009 Asociación Española de Pediatría. Published by Elsevier España, S.L. All rights reserved.

Introducción

Según los criterios diagnósticos del trastorno por déficit de atención e hiperactividad (TDAH) recogidos en el DSM (*Diagnostic and Statistical Manual of Mental Disorders* 'Manual Diagnóstico y Estadístico de los Trastornos Mentales')¹, para que a un sujeto se lo considere con TDAH ha de presentar 6 o más síntomas de inatención o hiperactividad-impulsividad en al menos 2 contextos durante, por lo menos, los 6 meses previos a la evaluación. No obstante, el DSM no menciona qué ítems es más efectivo combinar², ni establece las potenciales diferencias diagnósticas entre combinar unos y otros dentro de una misma escala. El presente trabajo se enmarca en esta línea de investigación que ya se ha abordado anteriormente³, especialmente en el caso de Power et al⁴.

En el anterior trabajo de los autores del presente artículo³ se replica la primera de las cuestiones planteadas y analizadas en el trabajo de Power et al⁴ (determinar el poder predictivo de cada uno de los síntomas diagnósticos del DSM-IV para el TDAH en las escalas de padres y de maestros en una muestra española); sin embargo, en el trabajo de Power se planteaban otras 2 cuestiones: ¿qué combinación de ítems es la óptima para establecer un diagnóstico a partir de los datos de cada informador? y ¿qué método de combinación de ítems es el mejor para predecir un diagnóstico de TDAH? El objetivo del presente trabajo se enmarca en ofrecer datos en relación con la primera de estas 2 cuestiones.

La hipótesis de partida fue que no todas las combinaciones de ítems tenían el mismo valor predictivo, y que algunas eran mejores a la hora de establecer o descartar un diagnóstico de TDAH. También es posible que estas combinaciones varíen en función de la evaluación de padres y maestros.

Pacientes y métodos

La muestra del presente trabajo deriva de un estudio poblacional prospectivo en la isla de Mallorca^{5,6}. En total, se evaluó a 1.509 niños y niñas a través de un muestreo estratificado y polietápico en una proporción 1:27. La muestra de niños, uniforme tanto en edad como en género, se evaluó con las ADHD RS (Attention Deficit Hyperactivity Disorder Rating Scales) -IV de DuPaul para maestros y padres⁷ para determinar la prevalencia del TDAH en la isla. En esta escala, cada ítem es coincidente con los criterios o síntomas diagnósticos del DSM-IV y todos se puntúan de 0 a 3: a mayor puntuación, mayores pruebas de problemas. La sospecha de TDAH se estableció a partir de puntuaciones superiores al percentil (P) 90, en la baremación original de la escala, en las subescalas de inatención e hiperactividad-impulsividad, tanto para padres como para maestros. La tasa global de prevalencia del TDAH combinado fue del 4,6%, y el hecho de que los baremos de la escala estén ajustados por edad y sexo sin duda favoreció a que no se detectaran diferencias en estas variables. No obstante, los cambios en las clasificaciones diagnósticas, así como el empleo de unos u otros criterios diagnósticos influirán notablemente en las tasas resultantes de los estudios epidemiológicos⁸⁻¹⁰.

En este trabajo, y de modo similar a los trabajos anteriores ya citados^{3,4}, se estudiaron las puntuaciones que proporcionaron tanto los padres como los maestros en la escala de inatención, de forma separada a las que otorgaron ambos evaluadores en la escala de hiperactividad-impulsividad. Para determinar el grado según el que una conducta habría de darse para que se la considere síntoma de un TDAH, se consideraron como ítems significativos (positivos) todos aquellos que hubieran recibido (de los diferentes evaluadores) una puntuación de 2 o superior. Por tanto,

ítems negativos son aquéllos con una puntuación igual o inferior a 1.

El siguiente paso fue calcular el poder predictivo total o valor predictivo total corregido (cTPV) mediante el índice de congruencia kappa¹¹ de cada uno de los síntomas (tanto para cada subescala como para cada tipo de evaluador). De este modo, se obtuvo un *ranking* de los síntomas (o ítems), en función de la proporción de casos de sospecha de TDAH que eran capaces de detectar correctamente.

A continuación se procedió a establecer diferentes combinaciones de ítems, de forma progresiva dentro de cada subtipo, en el orden en que éstos habían sido capaces de detectar el trastorno, con el fin de conocer cuáles de estas combinaciones eran las más importantes y mejores predictoras de un TDAH, y cuáles eran mejores para establecer el descarte.

Para determinar el número óptimo de síntomas de inatención de cada evaluador, primero se estudió la capacidad predictiva del ítem con cTPV más elevado. Posteriormente, se examinó la capacidad diagnóstica de los 2 ítems con mayor cTPV. A continuación, la capacidad de los 3 ítems de mayor cTPV, y así sucesivamente hasta incluir a todos. Seguidamente, se repitió el proceso con los ítems de hiperactividad-impulsividad.

De acuerdo con el enfoque de Power^{4,12}, para estudiar la capacidad predictiva de los ítems se utilizaron los siguientes indicadores para cada combinación de ítems: la sensibilidad, la especificidad, el poder predictivo positivo (PPP, y el mismo valor corregido por kappa [cPPP]) y el poder predictivo negativo (NPP, y el mismo valor corregido por kappa [cNPP]). Como criterios de inclusión (es decir, detección de sospecha de TDAH) se establecieron los siguientes puntos de corte: sensibilidad de 0,50 o superior y cPPP de 0,65 o superior. Los criterios de exclusión (es decir, capacidad de rechazar sospecha de TDAH) fueron los siguientes: especificidad igual o superior a 0,50 y cNPP superior o igual a 0,65.

Resultados

En la *tabla 1* se presenta el *ranking* de los ítems o criterios diagnósticos en función de la proporción de casos de sospecha de TDAH que han sido capaces de detectar correctamente; es decir, en función de los valores de cTPV. En la subescala de inatención, para los padres, destacan los ítems 4, 5 y 2 (todos por encima del 0,30). En el caso de los maestros, sólo el ítem 7 está por encima del 0,30, y los siguientes ítems son el 3, el 4 y el 5. En la subescala de hiperactividad-impulsividad tanto para los padres como para los maestros, los primeros ítems son el 4, el 2 y el 3. En el caso de los padres ninguno está por encima del 0,30 y, en el caso de los maestros, sólo el ítem 4.

A continuación, se aplicó el procedimiento descrito anteriormente para valorar la capacidad predictiva de las distintas combinaciones de ítems para detectar y descartar la sospecha de TDAH.

En el caso de la valoración de los padres en la escala de inatención (*tabla 2*), el uso de un solo ítem (el 4) («*no sigue instrucciones de principio a fin y no termina el trabajo asignado*») puede resultar suficiente para descartar la presencia de sospecha de TDAH, según los criterios de

Tabla 1 Valores del poder predictivo total corregido para *ranking* de padres y maestros de los ítems de trastorno por déficit de atención e hiperactividad

| Ítems | Padres | Maestros |
|---|--------|----------|
| IN-1: no logra prestar atención... | 0,28 | 0,21 |
| IN-2: tiene dificultad para mantener su atención... | 0,30 | 0,24 |
| IN-3: no parece escuchar... | 0,28 | 0,28 |
| IN-4: no sigue instrucciones... | 0,41 | 0,28 |
| IN-5: tiene dificultad para organizar.... | 0,40 | 0,28 |
| IN-6: evita tareas de esfuerzo mental... | 0,28 | 0,26 |
| IN-7: pierde cosas necesarias... | 0,29 | 0,31 |
| IN-8: se distrae fácilmente... | 0,16 | 0,18 |
| IN-9: es olvidadizo de las actividades diarias... | 0,28 | 0,25 |
| HI-1: mueve manos o pies constantemente... | 0,11 | 0,21 |
| HI-2: no permanece sentado en el aula... | 0,25 | 0,24 |
| HI-3: corre o se encarama de manera excesiva... | 0,24 | 0,29 |
| HI-4: dificultad para jugar tranquilamente... | 0,29 | 0,31 |
| HI-5: está siempre «en marcha»... | 0,11 | 0,21 |
| HI-6: habla excesivamente... | 0,10 | 0,14 |
| HI-7: responde impulsivamente... | 0,19 | 0,24 |
| HI-8: no puede esperar turno... | 0,23 | 0,28 |
| HI-9: interrumpe o se entromete... | 0,22 | 0,23 |

HI: subescala de hiperactividad-impulsividad; IN: subescala de inatención.

exclusión establecidos. Así pues, la probabilidad de que del grupo de «normales» un niño cualquiera dé negativo en este ítem es del 68% (con el control de los posibles efectos del azar). Por otra parte, la probabilidad de no presentar el trastorno cuando el ítem es negativo llega al 94%.

Por el contrario, para establecer la sospecha de TDAH no hubo ninguna combinación que superara los criterios de inclusión establecidos. La más elevada fue la combinación conjunta de los 2 ítems: «*no sigue instrucciones de principio a fin y no termina el trabajo asignado*» y «*tiene dificultad para organizar las tareas y las actividades que debe llevar a cabo*». La probabilidad de que, al resultar significativos estos 2 ítems, un sujeto presente sospecha de TDAH es del 46% (previamente se controlaron los factores provocados por el azar). Por otro lado, la probabilidad de que un niño cualquiera perteneciente al grupo de sospechosos de presentar un TDAH haya obtenido puntuaciones significativas en ambos ítems en la valoración parental es del 63%.

En la evaluación del maestro en la escala de inatención, también mediante un único ítem significativo, el 7 («*pierde cosas que son necesarias para llevar a cabo tareas o actividades*»), puede cumplirse el descarte. La probabilidad de que del grupo de «normales» un niño cualquiera dé negativo en este ítem es del 61% (con el control de factores de azar). Por otro lado, la probabilidad de no presentar sospecha de trastorno cuando se ha dado negativo en el ítem es del 92%.

Tabla 2 Estimaciones de utilidad de los síntomas para diagnosticar y descartar un trastorno por déficit de atención e hiperactividad, mediante el empleo de los síntomas de inatención mejor predictores

| Número del ítem | Tasa base | Sensibilidad | Especificidad | PPP | cPPP | NPP | cNPP |
|-----------------------------|-----------|--------------|---------------|------|------|------|------|
| Puntuaciones padres | | | | | | | |
| un solo ítem | 0,08 | 0,70 | 0,94 | 0,32 | 0,29 | 0,99 | 0,68 |
| 2 ítems | 0,05 | 0,63 | 0,97 | 0,48 | 0,46 | 0,99 | 0,61 |
| 3 ítems | 0,02 | 0,35 | 0,99 | 0,51 | 0,50 | 0,98 | 0,34 |
| 4 ítems | 0,02 | 0,31 | 0,99 | 0,65 | 0,64 | 0,98 | 0,30 |
| 5 ítems | 0,02 | 0,28 | 0,99 | 0,65 | 0,64 | 0,97 | 0,27 |
| 6 ítems | 0,01 | 0,19 | 0,99 | 0,63 | 0,61 | 0,97 | 0,18 |
| Puntuaciones maestro | | | | | | | |
| un solo ítem | 0,10 | 0,65 | 0,92 | 0,24 | 0,21 | 0,99 | 0,61 |
| 2 ítems | 0,05 | 0,44 | 0,97 | 0,32 | 0,30 | 0,98 | 0,42 |
| 3 ítems | 0,04 | 0,39 | 0,98 | 0,38 | 0,35 | 0,98 | 0,37 |
| 4 ítems | 0,03 | 0,35 | 0,98 | 0,38 | 0,36 | 0,98 | 0,33 |
| 5 ítems | 0,03 | 0,31 | 0,98 | 0,39 | 0,36 | 0,97 | 0,29 |
| 6 ítems | 0,03 | 0,26 | 0,98 | 0,34 | 0,32 | 0,97 | 0,24 |

Los ítems fueron combinados progresivamente en el orden en que eran capaces de detectar un trastorno por déficit de atención e hiperactividad de forma precisa (según el índice kappa).

cNPP: poder predictivo negativo corregido; cPPP: poder predictivo positivo corregido; NPP: poder predictivo negativo; PPP: poder predictivo positivo.

A la hora de establecer el diagnóstico, tampoco hubo ningún valor de cPPP que se aproximase siquiera al grado habitualmente considerado como aceptable (0,40). De nuevo, únicamente el ítem 7 de inatención («*pierde cosas que son necesarias para llevar a cabo tareas o actividades*») superó claramente el punto de corte de 0,50 propuesto para la sensibilidad. Por tanto, la probabilidad de que un niño que pertenece al grupo de sospecha de TDAH obtenga la significación en este ítem es del 65%. Sin embargo, la probabilidad de que se tenga sospecha de TDAH en un niño con un resultado positivo en el mismo ítem sólo es del 21%.

Al analizar la escala de hiperactividad-impulsividad que evaluaron los padres (tabla 3) en relación con el descarte, se pudo apreciar que la probabilidad de que un niño reciba una valoración no significativa del ítem 4 («*tiene dificultades para jugar tranquilamente o para llevar a cabo actividades en su tiempo libre de manera calmada*»), si se sabe que forma parte del grupo que no presenta sospecha de TDAH, es del 99%. Pero, por el contrario, la probabilidad de que un niño cualquiera dé negativo en este ítem y no presente sospecha de TDAH es de sólo el 31% (con control de los efectos del azar). Por otro lado, a la hora de establecer la sospecha de diagnóstico prácticamente ninguna combinación fue lo bastante fiable.

En el caso de la escala de hiperactividad-impulsividad que evaluaron los maestros, el mismo ítem 4 podría considerarse suficiente para descartar un TDAH. La probabilidad de que un niño del grupo sin sospecha de TDAH obtenga en ese ítem una puntuación negativa es del 95%. No obstante, la probabilidad de que un niño cualquiera con una puntuación negativa en este ítem no presente sospecha de TDAH es de, tan sólo, el 46% (con control de los factores de azar). En la sospecha de presencia del trastorno sucedió lo mismo que para los padres: ninguna combinación de ítems fue lo suficientemente fiable como para poder establecer sospechas de TDAH.

Al observar el cNPP y el cPPP, ambos se mantuvieron (en todos los casos) por debajo del 0,50; es decir, lejos de los valores exigidos.

Discusión

El grado predictivo de los criterios diagnósticos del TDAH aplicados en el DSM-IV y transformados en ítems de escalas para maestros y padres no es el mismo. Luego, en principio, es evidente que la línea de trabajo que intenta complementar el método politético propio del DSM-IV tiene interés. Para empezar, se destacarán algunas conclusiones generales.

En primer lugar, hay una tendencia muy acusada a que los análisis de las combinaciones de ítems (tanto para maestros como para padres y tanto para la subescala de inatención como para la de hiperactividad-impulsividad) den puntuaciones mucho más altas en especificidad (o NPP) que en sensibilidad (o PPP). Por tanto, la primera conclusión es que las combinaciones de ítems (sean cuales sean) son más robustas al detectar la proporción de sujetos «normales» que la de sujetos «clínicos» o, en otras palabras, la probabilidad de obtener falsos positivos es baja, pero no así la de obtener falsos negativos.

En segundo lugar, los puntos de corte, tanto para «inclusión» como para «exclusión», se han producido sólo en determinados valores (probablemente, como efecto de trabajar con tasas base muy bajas; véase más adelante en las limitaciones una ampliación de esta cuestión) y de manera muy variable. Por ejemplo, todas las combinaciones de ítems han presentado una especificidad extrema, pero sólo en 4 combinaciones se ha igualado o superado el 0,50 en sensibilidad (3 para la escala de inatención y una para la escala de hiperactividad-impulsividad). En los valores cPPP no se ha llegado al 0,65 en ninguna combinación y en los

Tabla 3 Estimaciones de utilidad de los síntomas para diagnosticar y descartar un trastorno por déficit de atención e hiperactividad, mediante el empleo de los síntomas de hiperactividad-impulsividad mejor predictores

| Número del ítem | Tasa base | Sensibilidad | Especificidad | PPP | cPPP | NPP | cNPP |
|-----------------------------|-----------|--------------|---------------|------|------|------|------|
| Puntuaciones padres | | | | | | | |
| un solo ítem | 0,04 | 0,33 | 0,97 | 0,31 | 0,28 | 0,98 | 0,31 |
| 2 ítems | 0,02 | 0,26 | 0,99 | 0,42 | 0,40 | 0,97 | 0,24 |
| 3 ítems | 0,01 | 0,19 | 0,99 | 0,48 | 0,46 | 0,97 | 0,17 |
| 4 ítems | 0,01 | 0,15 | 0,99 | 0,47 | 0,45 | 0,97 | 0,14 |
| 5 ítems | 0,01 | 0,11 | 0,99 | 0,46 | 0,44 | 0,97 | 0,10 |
| 6 ítems | 0,01 | 0,15 | 0,99 | 0,47 | 0,45 | 0,97 | 0,14 |
| Puntuaciones maestro | | | | | | | |
| un solo ítem | 0,07 | 0,50 | 0,95 | 0,26 | 0,23 | 0,98 | 0,46 |
| 2 ítems | 0,04 | 0,37 | 0,97 | 0,35 | 0,33 | 0,98 | 0,35 |
| 3 ítems | 0,03 | 0,35 | 0,98 | 0,39 | 0,37 | 0,98 | 0,33 |
| 4 ítems | 0,03 | 0,31 | 0,98 | 0,38 | 0,35 | 0,97 | 0,29 |
| 5 ítems | 0,03 | 0,30 | 0,98 | 0,37 | 0,35 | 0,97 | 0,28 |
| 6 ítems | 0,03 | 0,26 | 0,98 | 0,35 | 0,33 | 0,97 | 0,24 |

Los ítems fueron combinados progresivamente en el orden en que eran capaces de detectar un trastorno por déficit de atención e hiperactividad de forma precisa (según el índice kappa).

cNPP: poder predictivo negativo corregido; cPPP: poder predictivo positivo corregido; NPP: poder predictivo negativo; PPP: poder predictivo positivo.

valores cNPP sólo en una de inatención (aunque en los 2 indicadores ha habido hasta 6 combinaciones superiores al 0,60, que también sería un valor todavía respetable). Por tanto, los datos refuerzan la idea ya expuesta de mayor capacidad de detección de verdaderos negativos. Por otra parte, la tendencia general (al menos en el presente trabajo, en el que se detectaba sospecha de TDAH combinado a través del P90 de las subescalas de inatención e hiperactividad-impulsividad) es que las combinaciones en los ítems de inatención tienen algo más de valor predictivo que en los ítems de hiperactividad-impulsividad.

En tercer lugar, y en relación con los valores promedio de todas las combinaciones, puede decirse que la sensibilidad es algo mayor en los maestros que en los padres (especialmente cuando puntúan en la escala de hiperactividad-impulsividad); luego, los primeros son algo más fiables al detectar «verdaderos positivos». En cambio, prácticamente no habría diferencias entre los 2 tipos de evaluadores en especificidad. La tendencia, como suele ser normal en estos casos, se invierte en el caso del cPPP; ahora son los padres quienes presentan valores más altos y, por tanto, la probabilidad de presentar el trastorno sería más alta si diera positivo en padres que si diera en maestros (en términos promedio). En el cNPP sólo hay diferencias en hiperactividad-impulsividad, y el valor promedio de los maestros es superior al de los padres. Por tanto, la probabilidad de no presentar el trastorno sería teóricamente más alta si diera negativo en las combinaciones de estos ítems en la evaluación de los maestros, que en la de los padres.

Al adentrarse ya más específicamente en el objetivo principal del presente trabajo, lo que también parece evidente es que las combinaciones de ítems no tienen el mismo valor predictivo a la hora de proponer la sospecha-descarte de TDAH, como se había hipotetizado. Por tanto, el

uso de la combinación azarosa del DSM-IV es, como mínimo, discutible. Por ejemplo, en el caso de la escala de inatención, los ítems 4 y 5 para padres y el ítem 7 para los maestros tienen una capacidad predictiva (tanto de detección como de descarte) bastante superior a los demás ítems. En el caso de la escala de hiperactividad-impulsividad es más difícil señalar combinaciones concretas porque los valores generales son más bajos, pero, especialmente el ítem 4, en el caso de los maestros y también (aunque en menor medida) en el caso de los padres, ofrece grados de sensibilidad y de cNPP más altos que cualquier otra combinación, y de forma relativa frente a las otras combinaciones de 2 o más ítems; también sus especificidades y valores cPPP son más destacables. Por tanto, a la luz de los datos presentados, ¿qué ítem o combinación de ítems es la más fiable en la detección-descarte de TDAH? Pues los ítems 4 (*no sigue instrucciones...*) y 5 (*tiene dificultades para organizarse...*) de inatención de padres, el 7 (*pierde cosas necesarias...*) en inatención de maestros, y el ítem 4 (*dificultad para jugar tranquilamente...*) especialmente en maestros, pero también en padres.

Sin embargo, nuestros datos presentan algunas limitaciones que hay que tener presentes para futuros estudios y, por otra parte, las propias características del TDAH también aconsejan no perder de vista algunas consideraciones posteriores. En cuanto a las limitaciones, básicamente se destacan 2:

Uno de los principales problemas con los que cuenta este estudio a la hora de establecer una comparación con el estudio de Power et al es la tasa base, es decir, el porcentaje de niños del total de la muestra que puntuó positivo en el ítem o conjunto de ítems, tuvieran o no TDAH. Las diferencias encontradas entre ambos trabajos en este aspecto son debidas a que la muestra empleada en el actual artículo reunía tanto a sujetos con problemas de conducta o

aprendizaje como a controles, para reflejar las variaciones del TDAH en la población general escolar, y evitar los posibles sesgos que pudieran producirse en escenarios clínicos, como ya hicieron otros investigadores en el pasado^{13,14}. En cambio, en la investigación que llevaron a cabo Power et al, únicamente se trató con sujetos que se refirieron por problemas de conducta o aprendizaje, por lo que no es de extrañar que se encuentren puntuaciones más elevadas en el segundo caso, a la hora de predecir un diagnóstico de TDAH. No obstante, aunque numerosas investigaciones han apoyado la fiabilidad y validez de la ADHD RS-IV para la evaluación de los síntomas de inatención e hiperactividad-impulsividad^{15,16}, en 1984 Bergman y Pantell¹⁷ se dieron cuenta de que, en el caso de trastornos con una baja prevalencia, incluso las escalas de evaluación conductual con una excelente sensibilidad y especificidad podrían tener un valor cuestionable al usarse en la población general. Esto es debido a que la sensibilidad y la especificidad consideran únicamente la tasa de diagnósticos correctos. Este dato indica que se requieren futuros estudios que empleen muestras con una tasa base más elevada, que permita analizar la escala de forma más adecuada.

La segunda de las limitaciones para tener en cuenta en relación con el presente estudio, y que se pretende solventar de cara a futuros trabajos, es el hecho de haber empleado las escalas ADHD de DuPaul tanto para la selección de la muestra clínica, como para el análisis posterior de los ítems. A pesar de que la puntuación de corte requerida en todas las escalas y para todos los evaluadores fue alta (P90) de cara a descartar participantes dudosos las decisiones sobre la presencia o ausencia de un TDAH se basaron en el consenso de la información provista por padres y maestros¹⁸. Por tanto, sería conveniente establecer el diagnóstico de TDAH mediante una vía distinta y emplear la escala DuPaul en el análisis posterior de los ítems para comprobar cómo se comporta ésta realmente y, de esta forma, evitar posibles sesgos en los resultados de la investigación.

Por último, y en relación con las consideraciones que antes mencionábamos, hay que decir que una fortaleza de este estudio es que la muestra de sujetos fue uniforme, lo que la convierte en representativa de una población diversa de estudiantes, ya que investigaciones anteriores han mostrado importantes diferencias entre niños y niñas con respecto a los síntomas del TDAH^{19,20}. Que no se apreciaran diferencias significativas por razones de sexo, curso o zona conduce a cierto optimismo en la aplicación de los criterios diagnósticos a través de los sexos y las distintas edades, aunque es evidente que podría resultar necesario llevar a cabo análisis más minuciosos que pudieran servir para detectar diferencias de desarrollo en la aplicabilidad de los criterios diagnósticos. En este caso, sólo se ha abarcado el período entre los 6 y los 12 años. Por otra parte, no hay que perder de vista que las escalas de evaluación bien desarrolladas simplifican y estandarizan los datos, pero no son suficientes para emitir un diagnóstico²¹. Por esto, un diagnóstico preciso de TDAH requiere un enfoque multimétodo que incluya la obtención de información en diferentes contextos de todos los informadores significativos relacionados con el sujeto²², que permita tener una visión global de la problemática del niño. Aunque los resultados de las escalas de evaluación son elementos potencialmente valiosos del proceso de evaluación, y las guías recomiendan

encarecidamente el empleo de este tipo de escalas de banda estrecha específicas del TDAH^{23,24}, no pueden confundirse sus resultados con un diagnóstico definitivo de TDAH, y deben administrarse también entrevistas diagnósticas estructuradas, historias clínicas, registros, observaciones, medidas de laboratorio²⁵⁻³⁰, etc. Posteriormente, es necesario integrar la información en un proceso que aún no está del todo claro³¹.

En definitiva, el presente estudio muestra el valor y las limitaciones del empleo de las escalas en la evaluación diagnóstica del TDAH. Muchas escalas de evaluación se usan diariamente en el diagnóstico del TDAH, pero raramente estos instrumentos se han analizado para determinar su utilidad en el diagnóstico del TDAH. Es imprescindible, ante todo, saber cómo utilizarlas, y aún se conoce muy poco acerca de su utilidad para tomar decisiones diagnósticas³². Los datos de este estudio, así como otros desarrollados por el grupo de Power, muestran que la lista de ítems-criterios diagnósticos del DSM-IV no son equiparables en su poder predictivo, que puede ser factible en el futuro presentarlos de modo jerarquizado, e incluso que pueden desarrollarse instrumentos de apoyo más eficaces (de menos tiempo y menos coste). Para esto deberán redoblar los estudios en esta línea de trabajo, puesto que los datos todavía son escasos.

Agradecimientos

El presente trabajo se ha realizado en parte gracias al proyecto de investigación SEJ2004-01363 del Ministerio de Educación y Ciencia, Gobierno de España.

Bibliografía

1. American Psychiatric Association. Manual diagnóstico y estadístico de los trastornos mentales-IV-Texto revisado (DSM IV-TR). Barcelona: Masson; 2001.
2. Frick PJ, Lahey BB, Applegate B, Kerdyck L, Ollendick T, Hynd GW, et al. DSM-IV field trials for the disruptive behavior disorders: Symptom utility estimates. *J Am Acad Child Adolesc Psychiatry*. 1994;33:529-39.
3. Cardo E, Bustillo M, Servera M. Valor predictivo de los criterios del DSM-IV en el diagnóstico del TDAH y sus diferencias culturales. *Rev Neurol*. 2007;44:S19-22.
4. Power TJ, Costigan TE, Leff SS, Eiraldi RB, Landau S. Assessing ADHD across settings: Contributions of behavioral assessment to categorical decision making. *J Clin Child Psychol*. 2001;30:399-412.
5. Cardo E, Mateo M. Prevalencia del trastorno de déficit de atención e hiperactividad. *Rev Neurol*. 2005;40:S11-5.
6. Cardo E, Servera M, Llobera J. Estimación de la prevalencia del trastorno por déficit de atención e hiperactividad en población normal de la isla de Mallorca. *Rev Neurol*. 2007;44:10-4.
7. DuPaul GJ, Anastopoulos AD, Power TJ, Reid R, McGoey KE, Ikeda MJ. Parent ratings of ADHD symptoms: Factor structure, normative data, and psychometric properties. *J Psychopathol Behav Assess*. 1998;20:83-102.
8. Montiel-Nava C, Peña JA, López M, Salas M, Zurza JR, Montiel-Barbero I, et al. Estimaciones de la prevalencia del trastorno por déficit de atención-hiperactividad en niños marabinos. *Rev Neurol*. 2002;35:1019-24.
9. Cornejo JW, Osío O, Sánchez Y, Carrizosa J, Sánchez G, Grisales H, et al. Prevalencia del trastorno por déficit de

- atención-hiperactividad en niños y adolescentes colombianos. *Rev Neurol.* 2005;40:716–22.
10. Pineda DA, Lopera F, Henao GC, Palacio JD, Castellanos FX, Grupo de Investigación Fundema. Confirmación de la alta prevalencia del trastorno por déficit de atención en una comunidad colombiana. *Rev Neurol.* 2001;32:217–22.
 11. Chen WJ, Faraone SV, Biederman J, Tsuang MT. Diagnostic accuracy of the Child Behavior Checklist scales for attention-deficit hyperactivity disorder: A receiver-operating characteristic analysis. *J Consult Clin Psychol.* 1994;62:1017–25.
 12. Power TJ, Andrews TA, Eiraldi RB, Doherty BJ, Ikeda MJ, DuPaul GJ, et al. Evaluating attention deficit hyperactivity disorder using multiple informants: The incremental utility of combining teacher with parent reports. *Psychol Assess.* 1998;10:250–60.
 13. Jensen PS, Martin D, Cantwell DP. Comorbidity in ADHD: Implications for research, practice, and DSM-V. *J Am Acad Child Adolesc Psychiatry.* 1997;36:1065–79.
 14. Fergusson DM, Horwood J, Lynskey MT. Prevalence and comorbidity of DSM-III-R diagnoses in a birth cohort of 15 years olds. *J Am Acad Child Adolesc Psychiatry.* 1993;32:1127–34.
 15. DuPaul GJ, Power TJ, Anastopoulos AD, Reid R, McGoey K, Ikeda MJ. Teacher ratings of attention deficit hyperactivity disorder symptoms: Factor structure and normative data. *Psychol Assess.* 1997;9:436–44.
 16. DuPaul GJ, Power TJ, McGoey KE, Ikeda MJ, Anastopoulos AD. Reliability and validity of parent and teacher ratings of attention-deficit/hyperactivity disorder symptoms. *J Psychoed Assess.* 1998;16:55–68.
 17. Bergman DA, Pantell RH. The art and science of medical decision making. *J Pediatr.* 1984;104:649–56.
 18. Mitsis EM, McKay KE, Schulz KP, Newcorn JH, Halperin JM. Parent and teacher concordance for DSM-IV Attention-Deficit/Hyperactivity Disorder in a child-referred sample. *J Am Acad Child Adolesc Psychiatry.* 2000;39:308–13.
 19. Arnold LE. Sex differences in AD/HD: Conference summary. *J Abnorm Child Psychol.* 1996;24:555–70.
 20. Valdizán JR, Mercado E, Mercado-Undanivia A. Características y variabilidad clínica del trastorno por déficit de atención/hiperactividad en niñas. *Rev Neurol.* 2007;44:527–30.
 21. McGough JJ, McCracken JT. Assessment of attention deficit hyperactivity disorder: A review of recent literature. *Curr Opin Pediatr.* 2000;12:319–24.
 22. Barkley RA. Attention deficit hyperactivity disorder: A handbook for diagnosis and treatment. New York: Guilford; 1990.
 23. Brown RT, Freeman WS, Perrin JM, Stein MT, Amler RW, Feldman HM, et al. Prevalence and assessment of attention deficit/hyperactivity disorder in primary care settings. *Pediatrics.* 2001;107:E43.
 24. American Academy of Pediatrics, Committee on Quality Improvement, Subcommittee on Attention-Deficit/Hyperactivity Disorder. Clinical practice guideline: Diagnosis and evaluation of the child with attention-deficit/hyperactivity disorder. *Pediatrics.* 2000;105:1158–69.
 25. Atkins MS, Pelham WE. School-based assessment of attention deficit-hyperactivity disorder. En: Shaywitz SE, Shaywitz BA, editores. Attention deficit disorder comes of age: Toward the twenty-first century. Austin, TX: Pro-Ed; 1992. p. 69–88.
 26. Schaugency EA, Rothlind J. Assessment and classification of attention deficit hyperactivity disorders. *School Psych Rev.* 1991;20:187–202.
 27. Collett BT, Ohan JL, Myers KM. Ten-year review of rating scales. V: Scales assessing attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry.* 2003;42:1015–37.
 28. Wender E. ADHD symptoms and parent-teacher agreement. *J Dev Behav Pediatr.* 2004;25:48–9.
 29. DuPaul GJ. Assessment of ADHD symptoms: Comment on Gómez et al. *Psychol Assess.* 2003;15:113–7.
 30. Tripp G, Schaugency EA, Clarke B. Parent and teacher rating scales in the evaluation of attention-deficit hyperactivity disorder: Contribution to diagnosis and differential diagnosis in clinically referred children. *J Dev Behav Pediatr.* 2006;27:209–18.
 31. Wolraich ML. Author's response to commentaries. *J Dev Behav Pediatr.* 2004;25:52.
 32. Reid R, Maag JW. How many fidgets in a pretty much: A critique of behavior rating scales for identifying students with ADHD. *J Sch Psychol.* 1994;32:339–54.