

El análisis de curvas ROC en estudios epidemiológicos de psicopatología infantil: aplicación al cuestionario CBCL*

José Blas Navarro
Josep M. Doménech
Núria de la Osa
Lourdes Ezpeleta
Universitat Autònoma de Barcelona

Mediante el análisis de curvas ROC se estudia la precisión diagnóstica del Child Behavior Checklist (CBCL) y se obtiene el punto de corte óptimo en una muestra de 196 niños y adolescentes procedentes de centros de consulta pediátrica y psiquiátrica. Se utilizaron como patrones de referencia el grupo de procedencia, el diagnóstico según la entrevista diagnóstica estructurada DICA-R y el diagnóstico del clínico. Los resultados indican que la capacidad del CBCL para discriminar entre sujetos con y sin psicopatología depende en gran medida del patrón de referencia utilizado, siendo mejor el rendimiento de la prueba cuando se contrasta con el grupo de procedencia o con la entrevista diagnóstica estructurada. Como prueba de cribado, el punto de corte de la puntuación total del CBCL se situaría entre los valores 50 y 54 para optimizar la sensibilidad.

Palabras clave: Curvas ROC, CBCL, precisión diagnóstica, tests de screening.

ROC analysis was applied in order to study diagnostic accuracy of the Child Behavior Checklist (CBCL) and to obtain the optimal cut-off in a sample of 196 pediatric and psychiatric patients 6 to 17 years old. The group of origin, the diagnosis in the Diagnostic Interview for Children and Adolescents-Revised (DICA-R) and the clinician's diagnosis were used as external validators. The results indicate that the discriminant power of the CBCL for the presence or absence of psychopathology depends on the external validator used. The best results were obtained when it was considered the group of origin and the DICA-R diagnoses. As screening test, a cut-off between 50 and 54 gave the best sensitivity.

Keywords: ROC, CBCL, Diagnostic performance, Screening Tests.

* Este trabajo ha sido realizado gracias a las ayudas DGICYT PM91-209 y PM95-126 del Ministerio de Educación y Ciencia

Correspondencia: J.B. Navarro y J. M. Doménech, Departament de Psicobiologia i Metodologia de les Ciències de la Salut; N. de la Osa y L. Ezpeleta, Departament de Psicologia de la Salut i de Psicologia Social, Facultat de Psicologia, Universitat Autònoma de Barcelona. 08193 Bellaterra (Barcelona).

Los estudios epidemiológicos en psicopatología infantil utilizan frecuentemente pruebas de cribado (*screening*) que proporcionan un diagnóstico binario del estado del sujeto: sano o patológico, aplicando un punto de corte al resultado cuantitativo que ofrece el algoritmo de corrección de la prueba.

La precisión diagnóstica de una prueba de cribado se calcula a partir del número de sujetos correctamente clasificados por la prueba según un patrón de referencia (*Gold standard*) que indica el estado real del sujeto. El establecimiento de un patrón de referencia adecuado es uno de los más graves problemas con que se encuentra la investigación epidemiológica en psicopatología infantil, ya que no existe una «verdad absoluta» sobre la presencia o ausencia de patología mental, sino que ésta depende del criterio elegido en cada ocasión. El resultado de una prueba consolidada y consensuada por expertos se ha convertido en el referente más utilizado en este tipo de investigaciones, con la consiguiente paradoja de que una nueva prueba nunca podrá ser considerada mejor que la antigua, aunque aproxime mejor el estado real de los sujetos (Fletcher, 1988).

Una solución razonable a esta delicada situación consiste en evaluar la precisión diagnóstica a partir de varios patrones de referencia, obtenidos de forma independiente, pero evaluados con una metodología común que permita su comparación (Verhulst y Koot, 1992).

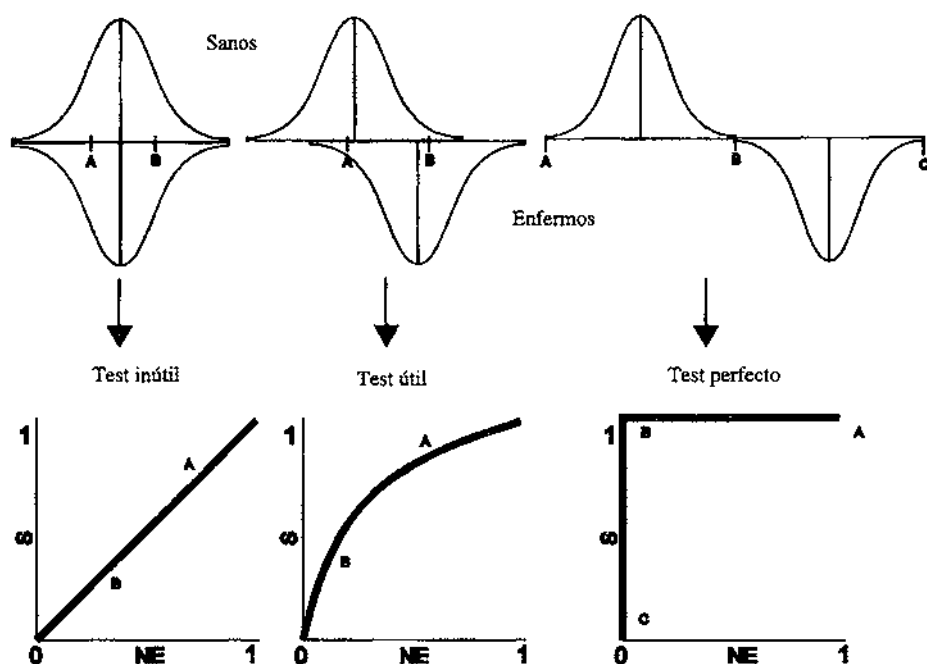
De entre los índices clásicos de evaluación de la precisión diagnóstica los más importantes son sensibilidad (S) y especificidad (E), y sus complementarios insensibilidad ($NS=1-S$) e inespecificidad ($NE=1-E$), que se calculan a partir de una tabla 2-2 en la que se cruzan, en filas, el diagnóstico binario obtenido al aplicar un punto de corte al resultado cuantitativo de la prueba y, en columnas, el patrón de referencia que indica el estado real del sujeto.

Sin embargo, estos índices presentan el inconveniente de depender del punto de corte aplicado a la prueba (Turner, 1978): un punto de corte bajo disminuye los falsos negativos a costa de incrementar la cantidad de falsos positivos, mientras que un punto de corte alto produce el efecto contrario. También existen otros índices como la eficacia, la ganancia de certidumbre, el índice de acuerdo kappa o el de asociación phi, pero todos ellos dependen tanto del punto de corte como de la prevalencia de la patología analizada en la población objeto de estudio.

El análisis de curvas ROC (*Receiver Operating Characteristics*) proporciona una medida global de precisión diagnóstica, independiente del punto de corte y de la prevalencia, de gran interés para muchos estudios epidemiológicos en psicopatología infantil, ya que permite comparar la eficacia de diferentes pruebas diagnósticas aplicadas a una misma muestra o de una misma prueba aplicada a diferentes muestras (Fombonne, 1991).

La curva ROC se obtiene representando la sensibilidad (porcentaje de verdaderos positivos), en el eje de ordenadas, y la inespecificidad (porcentaje de falsos positivos), en el eje de abscisas, para diferentes puntos de corte aplicados sobre el resultado cuantitativo de un test. Así pues, la curva ROC es independiente del punto de corte y de la prevalencia, cumpliendo así las dos condiciones enunciadas por Swets (1988) como imprescindibles en un indicador de precisión diagnóstica.

La curva ROC refleja el grado de solapamiento de las puntuaciones del test en los grupos de sujetos sanos y patológicos (véase Figura 1). Cuando el solapamiento es total (test inútil) la curva ROC recorre la diagonal positiva del gráfico, ya que para cualquier punto de corte sensibilidad (S) es igual a inespecificidad (NE). Cuando el solapamiento es nulo (test perfecto), la curva ROC recorre los bordes izquierdo y superior del gráfico, ya que para cualquier punto o bien $S=1$ o bien $NE=0$, existiendo algún punto de corte en que ambos $S=1$ y $NE=0$. En la práctica el solapamiento de puntuaciones entre sanos y enfermos será parcial, generando curvas ROC intermedias entre las dos situaciones planteadas (Weinstein y Fineberg, 1980).



A, B, C: Puntos de corte; S: Sensibilidad; NE: Inespecificidad

Figura 1. Curvas ROC según el solapamiento de las puntuaciones del test entre sanos y enfermos.

El área (W) bajo toda la curva (Simpson y Fitter, 1973) se ha convertido en el índice de resumen de una curva ROC más utilizado. El área bajo la curva de un test inútil es 0.5, reflejando que al emplearlo como prueba diagnóstica sólo clasifica correctamente un 50% de individuos, por el contrario, el área bajo la curva de un test perfecto es 1, ya que permite clasificar sin error el 100% de sujetos.

El análisis de curvas ROC proporciona 4 criterios de selección del punto de corte óptimo en función del indicador epidemiológico que se maximice (Green

y Swets, 1966; Egan, 1975): (I) una combinación ponderada de sensibilidad y especificidad, (II) el porcentaje total de sujetos correctamente clasificados, (III) la sensibilidad fijada la especificidad o (IV) la especificidad fijada la sensibilidad. La selección de uno u otro criterio se realizará teniendo en cuenta la prevalencia de la patología y la gravedad de los falsos negativos y los falsos positivos. En la Tabla 1 se presentan el criterio más adecuado en función de estas dos características y el punto de corte resultante.

TABLA 1. CRITERIO DE SELECCIÓN DEL PUNTO DE CORTE ÓPTIMO SEGÚN LA PREVALENCIA Y LAS CONSECUENCIAS DE LOS ERRORES DE CLASIFICACIÓN

Prevalencia	Gravedad de las malas clasificaciones	Criterios recomendados	Punto de corte resultante
Alta	FN > FP	IV, II	Muy bajo
	FN $\approx$ FP	II	Bajo
	FN < FP	III	Medio-Alto
Media	FN > FP	IV, II	Bajo
	FN $\approx$ FP	I, II	Medio
	FN < FP	III, II	Alto
Alta	FN > FP	IV	Bajo-Medio
	FN $\approx$ FP	II	Alto
	FN < FP	III, II	Muy alto

FN: Falsos negativos; FP: Falsos positivos

En este trabajo aplicamos el análisis de curvas ROC al *Child Behavior Checklist* (CBCL; Achenbach, 1991) con un doble objetivo: (1) estudiar su precisión diagnóstica mediante análisis ROC con diferentes patrones de referencia y (2) obtener el punto de corte óptimo para utilizar la prueba como instrumento de cribado.

Método

Diseño

La puntuación obtenida con el CBCL ha sido analizada utilizando tres criterios de clasificación distintos que han sido obtenidos de forma independiente en una misma muestra de sujetos. Según Verhulst y Koot (1992) los tres criterios empleados son los más habituales en psicopatología infantil.

– Criterio 1: Grupo de procedencia. Los sujetos de la muestra procedían de consulta pediátrica (no patológicos) o consulta psiquiátrica (patológicos).

– Criterio 2: Clasificación según una entrevista diagnóstica estructurada. La entrevista DICA-R pudo aplicarse al 95% de la muestra, ya que en 4 sujetos de procedencia pediátrica y 6 de psiquiátrica los padres no acudieron a la entrevista. Con este criterio un sujeto fue considerado patológico si presentaba uno o más diagnósticos positivos. En caso de no presentar ningún diagnóstico positivo el sujeto fue asignado al grupo de no patología.

– Criterio 3: Clasificación según criterio clínico. En el 92.4% de sujetos de procedencia psiquiátrica (67.9% de la muestra total) se tuvo acceso al diagnóstico DSM-III-R del clínico al que acudían a consulta. En el 7.6% restante no se obtuvo esta información porque el clínico no cumplimentó la lista de síntomas *Checklist*. Con este criterio un sujeto fue considerado patológico cuando tenía presente uno o más trastornos de los 25 evaluados por los clínicos (mediante la *Checklist DSM-III-R*).

Sujetos

La muestra total está compuesta por 196 niños, de edades comprendidas entre 6 y 17 años que acudían a consulta externa pediátrica o psiquiátrica en centros asistenciales de Barcelona.

La distribución por sexo y edad según los tres criterios de clasificación establecidos viene dada en la Tabla 2.

El nivel socioeconómico de la muestra total, según el índice de posición social de Hollingshead (1975) fue: clase I: 3.7%; II: 8.5%, III: 12.6%, IV: 33.2% y V: 42% (la clase I representa el nivel sociocultural más alto). En las submuestras derivadas de los criterios de clasificación 2 y 3 se obtuvieron distribuciones similares.

TABLA 2. DISTRIBUCIÓN DE SEXO Y EDAD EN LAS SUBMUESTRAS ASOCIADAS A LOS TRES PATRONES DE REFERENCIA

		Número de sujetos	Sexo		Edad Media (DE)
			Niños	Niñas	
Grupo de procedencia	No patológicos	52 (26.5%)	28 (53.8%)	24 (46.2%)	9.9 (2.3)
	Patológicos	144 (73.5%)	69 (47.9%)	75 (52.1%)	12.3 (3.1)
Grupo según DICA-R	No patológicos	59 (31.7%)	29 (49.2%)	30 (50.8%)	10.9 (2.9)
	Patológicos	127 (68.3%)	61 (48.0%)	66 (52.0%)	12.2 (3.1)
Grupo según criterio clínico	No patológicos	28 (21.1%)	12 (42.9%)	16 (57.1%)	11.5 (3.2)
	Patológicos	105 (78.9%)	53 (50.5%)	52 (49.5%)	12.3 (3.1)

DE: Desviación estándar

Material

— El *Child Behavior Checklist* (CBCL; Achenbach, 1991) es el cuestionario de evaluación de psicopatología infantil más utilizado con fines de cribado. Para el presente trabajo se utilizó la parte de *problemas de conducta* del cuestionario para padres, que está formada por 118 ítems que valoran, en una escala de 3 puntos, la frecuencia de aparición de determinados síntomas conductuales y emocionales. El CBCL da puntuaciones cuantitativas de 10 dimensiones psicopatológicas y una puntuación total que refleja la psicopatología global del sujeto. Dichas puntuaciones son transformadas a puntuaciones T, según baremos diferenciados por sexo y grupo de edad (4-11, 12-18 años), con el objetivo de eliminar la diferencia en presencia de patología de estos grupos. En el presente estudio se han utilizado los baremos proporcionados por Achenbach (1991). La fiabilidad y validez del CBCL ha sido ampliamente estudiada obteniéndose excelentes resultados (Achenbach, 1991).

— *Diagnostic Interview for Children and Adolescent-Revised* (DICA-R) (Reich, Shaika y Taibleson, 1991). La DICA-R es una entrevista diagnóstica estructurada de evaluación de psicopatología en niños y adolescentes de 6 a 17 años, que explora la presencia/ausencia, según criterios DSM-III-R, de 25 trastornos psicológicos que suelen presentarse en estas edades.

La DICA-R tiene tres formas, una para ser contestada por niños (6-12 años), otra por adolescentes (13-17 años) y una tercera por padres. Puesto que el CBCL lo contestan los padres, para esta entrevista hemos utilizado el mismo informante (DICA-R-P).

En la adaptación española se ha obtenido una fiabilidad entre entrevistadores y test-retest entre buena y excelente (Ezpeleta, de la Osa, Doménech, Navarro y Losilla, 1995; de la Osa, Ezpeleta, Doménech, Navarro y Losilla, 1996a). La validez convergente y discriminante fue aceptable (Ezpeleta, de la Osa, Júdez, Doménech, Navarro & Losilla, en prensa; de la Osa, Ezpeleta, Doménech, Navarro y Losilla, 1996b).

— *Checklist DSM-III-R* (Ezpeleta, 1994). Consiste en una lista de síntomas con dos opciones de respuesta (falso/verdadero) que evalúan la presencia/ausencia de los 25 trastornos DSM-III-R incluidos en la entrevista DICA-R. La *Checklist* fue rellenada por el clínico (psicólogos y psiquiatras) responsable del sujeto en estudio.

Procedimiento

En todos los casos de la muestra psiquiátrica los clínicos tenían el primer contacto con el niño y su familia. Los casos con sospecha de deficiencia mental fueron excluidos. Profesores y estudiantes de segundo ciclo de psicología, previamente entrenados en el uso de la entrevista y que desconocían el diagnóstico del clínico, administraban la DICA-R a los padres. A continuación se solicitaba al clínico responsable que rellenara la *Checklist* teniendo presente toda la información que considerara oportuna. En una segunda reunión con los padres se admi-

nistraban diferentes escalas y cuestionarios, entre los que se hallaba el CBCL.

En la muestra pediátrica se excluyeron los casos con enfermedades psicosomáticas o crónicas. Se aplicó la entrevista DICA-R y el cuestionario CBCL siguiendo un procedimiento muy similar al de la muestra psiquiátrica. Los sujetos de este grupo no fueron diagnosticados por un clínico.

El análisis del CBCL se ha realizado a partir de la puntuación total T y se ha estructurado en tres apartados:

- Análisis estadístico clásico: prueba *t* de Student de comparación de medias.

- Análisis de curvas ROC según los tres patrones de referencia establecidos: las curvas fueron obtenidas mediante aproximación no paramétrica (Hanley y McNeil, 1982, 1983).

- Selección del punto de corte óptimo: con el criterio de maximizar la especificidad fijada la sensibilidad.

Resultados

Análisis estadístico clásico

La Tabla 3 presenta el resultado de la comparación de medias (prueba *t* de Student) de la puntuación total T del CBCL entre los grupos de sujetos patológicos y no patológicos según los tres criterios de clasificación. Los sujetos patológicos puntuaron más alto que los no patológicos, siendo la diferencia estadísticamente significativa cuando el patrón de referencia fue el grupo de procedencia (criterio 1) o la entrevista DICA-R (criterio 2).

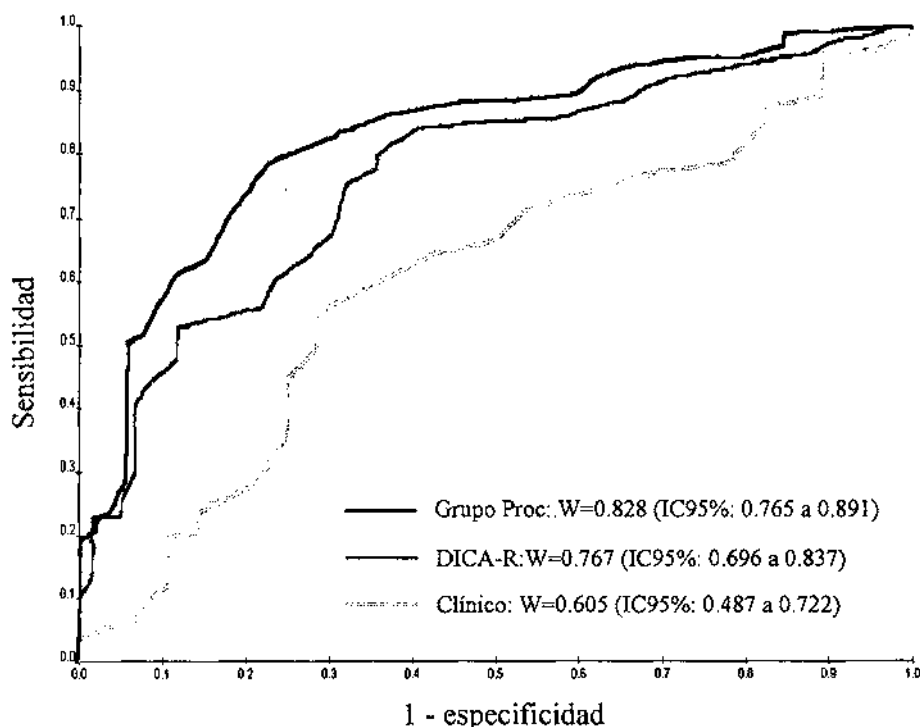
TABLA 3. COMPARACIÓN DE MEDIAS DE LA PUNTUACIÓN TOTAL T DEL CBCL ENTRE PATOLÓGICOS Y NO PATOLÓGICOS SEGÚN LOS TRES PATRONES DE REFERENCIA

		Punt. Total T Media (DE)	Diferencia medias (IC95%)	Signif.
Grupo de procedencia	No patológicos	49.4 (8.8)	12.3 (9.3 a 15.3)	< 0.001
	Patológicos	61.7 (9.8)		
Grupo según DICA-R	No patológicos	51.6 (9.6)	10.2 (7.1 a 13.3)	< 0.001
	Patológicos	61.8 (10.3)		
Grupo según criterio clínico	No patológicos	58.8 (9.5)	3.3 (-0.7 a 7.3)	0.101
	Patológicos	62.1 (9.7)		

DE: Desviación estándar, IC95%: Intervalo de confianza del 95%

Análisis de curvas ROC

Las curvas ROC no paramétricas de la puntuación total T del CBCL según los tres patrones de referencia, junto a las estimaciones del área total bajo la curva y del intervalo de confianza se representan en la Figura 2. Cuando el grupo de procedencia fue seleccionado como patrón de referencia la precisión diagnóstica del CBCL fue buena ($W=0.828$; IC95%: 0.765 a 0.891). La precisión diagnóstica disminuyó cuando el patrón de referencia fue la entrevista DICA-R ($W=0.767$; IC95%: 0.696 a 0.837). Cuando se consideró el criterio del clínico los resultados fueron pobres; el valor del área bajo la curva $W=0.605$ (IC95%: 0.487 a 0.722) indica que el CBCL carece de poder predictivo ya que su IC95% incluye el valor 0.5.



W: Área bajo la curva, IC95%: Intervalo de confianza del 95% de W

Figura 2. Curvas ROC de la puntuación total T del CBCL según los tres patrones de referencia.

La Tabla 4 presenta los valores de sensibilidad y especificidad y sus respectivos intervalos de confianza para los puntos de corte propuestos por diferentes autores (Achenbach, 1991; Fombonne, 1991). Los valores de sensibilidad y

TABLA 4. SENSIBILIDAD Y ESPECIFICIDAD DE LA PUNTUACIÓN TOTAL T DEL CBCL PARA DIFERENTES PUNTOS DE CORTE SEGÚN LOS TRES PATRONES DE REFERENCIA

P.C.	Grupo procedencia	DICA-R	Criterio clínico
50 S (IC95%) E (IC95%)	.88 (.83 a .92) .54 (.41 a .66)	.86 (.80 a .91) .42 (.31 a .54)	.88 (.82 a .93) .14 (.05 a .30)
53 S (IC95%) E (IC95%)	.80 (.73 a .85) .75 (.63 a .84)	.79 (.73 a .85) .64 (.53 a .75)	.80 (.72 a .86) .21 (.10 a .38)
55 S (IC95%) E (IC95%)	.76 (.69 a .81) .79 (.67 a .88)	.76 (.68 a .82) .68 (.56 a .78)	.77 (.69 a .84) .32 (.18 a .49)
57 S (IC95%) E (IC95%)	.64 (.57 a .71) .85 (.74 a .92)	.64 (.56 a .71) .73 (.62 a .82)	.67 (.58 a .74) .50 (.33 a .67)
60 S (IC95%) E (IC95%)	.51 (.44 a .58) .92 (.83 a .97)	.53 (.46 a .61) .86 (.77 a .93)	.55 (.47 a .63) .71 (.54 a .85)
63 S (IC95%) E (IC95%)	.46 (.39 a .54) .94 (.85 a .98)	.48 (.40 a .56) .88 (.79 a .94)	.49 (.41 a .58) .71 (.54 a .85)
65 S (IC95%) E (IC95%)	.37 (.31 a .45) .94 (.85 a .98)	.40 (.33 a .48) .93 (.85 a .98)	.39 (.31 a .47) .75 (.58 a .88)
67 S (IC95%) E (IC95%)	.29 (.23 a .36) .94 (.85 a .98)	.31 (.24 a .38) .93 (.85 a .98)	.28 (.21 a .37) .78 (.62 a .90)
70 S (IC95%) E (IC95%)	.21 (.15 a .27) .98 (.91 a .99)	.23 (.17 a .30) .98 (.92 a .99)	.21 (.15 a .29) .86 (.70 a .95)

P.C.: Punto de corte, S: Sensibilidad, E: Especificidad, IC95%: Intervalo de confianza del 95%

especificidad y sus intervalos de confianza reflejan que, con los puntos de corte habituales en la prueba, el CBCL es más específico que sensible.

Selección del punto de corte óptimo

Se calcularon los puntos de corte óptimos de la puntuación total T del CBCL según los dos criterios de clasificación con los que se obtuvieron valores de precisión diagnóstica que sugerían el uso del cuestionario como prueba de cribado: grupo de procedencia y entrevista DICA-R. No se ha considerado el diagnóstico clínico por la baja precisión del CBCL asociada a este criterio.

El uso del CBCL como instrumento de cribado comporta detectar un elevado porcentaje de casos patológicos (sensibilidad alta). Los falsos positivos deberán ser localizados en una segunda fase del estudio. En consecuencia, se ha calculado el punto de corte óptimo según dos criterios: (1) fijar la sensibilidad

entre 0.75 y 0.85 y maximizar la especificidad, y (2) fijar la sensibilidad entre 0.85 y 0.95 y maximizar la especificidad.

En la Tabla 5 se presentan los puntos de corte obtenidos con cada criterio junto a los valores de sensibilidad y especificidad correspondientes y sus intervalos de confianza.

TABLA 5. PUNTOS DE CORTE DE LA PUNTUACIÓN TOTAL T DEL CBCL

		P.C.	S (IC95%)	E (IC95%)
Sensibilidad entre 0.75-0.85	Grupo procedencia	54	.79 (.73 a .84)	.77 (.65 a .86)
	DICA-R	53	.79 (.73 a .85)	.64 (.53 a .75)
Sensibilidad entre 0.85-0.95	Grupo procedencia	50	.88 (.83 a .92)	.54 (.41 a .66)
	DICA-R	46	.92 (.87 a .96)	.29 (.19 a .40)

P.C.: Punto de corte, S: Sensibilidad, E: Especificidad, IC95%: Intervalo de confianza del 95%

Discusión

Desde sus inicios en la década de los 50, el análisis de curvas ROC se ha consolidado como el método más adecuado para contrastar la capacidad diagnóstica de un test con resultado cuantitativo. De entre sus numerosas propiedades cabe destacar que la medida de precisión diagnóstica que proporciona es independiente de la prevalencia y del punto de corte, permitiendo así la comparación directa de resultados de diferentes estudios de precisión realizados sobre una misma prueba diagnóstica. El análisis de curvas ROC no está exento de defectos, si bien su repercusión práctica es mínima. Uno de los inconvenientes es que el valor de los puntos de corte y el número de sujetos sanos y patológicos no aparecen en el gráfico.

En el terreno concreto de la epidemiología en psicopatología infantil y de adultos surge, como problema añadido a la evaluación de tests, la dificultad de encontrar un patrón de referencia que refleje sin error el estado real del sujeto (Lane, Pollard y Cox, 1990; Robins, 1985; Vaillant y Schnurr, 1988). Una de las soluciones a este problema consiste en el establecimiento de varios patrones de referencia. El grupo de procedencia del sujeto, la clasificación obtenida por pruebas diagnósticas ampliamente validadas y el criterio de un experto constituyen algunas de las alternativas más utilizadas.

Respecto al CBCL, los análisis realizados reflejan una buena capacidad diagnóstica para distinguir entre sujetos patológicos y no patológicos en población de procedencia mixta (pediátrica y psiquiátrica). Sin embargo, cuando su poder diagnóstico se pone a prueba en población de procedencia exclusivamente psiquiátrica (criterio 3) se obtiene sólo un 60.5% de sujetos correctamente clasificados, que desaconseja su uso aplicado en sujetos de estas características. Ello

puede ser debido a dos motivos: o bien el clínico no contempla todos los problemas conductuales (síntomas) como el principal determinante de psicopatología, o bien el CBCL no es capaz de discriminar entre sujetos con un grado de sintomatología basal elevado. En efecto, la muestra de procedencia psiquiátrica estaba formada por niños que tenían algún problema y por ello pedían ayuda al especialista. En este sentido, esta muestra era muy homogénea en cuanto a presencia de síntomas, aunque no necesariamente agrupados en un síndrome (patológico-no patológico).

La falta de concordancia (entre informantes, instrumentos, clínicos, etc.) en el ámbito de la psicopatología tanto infantil como de adultos es la regla más que la excepción (Godoy, 1996). En este caso los instrumentos empleados así como su trasfondo teórico varían: por un lado el instrumento de cribado es un cuestionario sintomatológico del que se ha derivado un sistema de clasificación psicopatológico dimensional (registra intensidad del trastorno), por el otro, la referencia es el juicio del clínico siguiendo un sistema de clasificación categorial (registra presencia o ausencia del trastorno). El debate actual sobre cuál de los dos sistemas es más adecuado está abierto. Aunque los psicólogos son más partidarios de los sistemas dimensionales porque, frente a los categoriales, son más precisos y evitan que se pierda información, el consenso está en el uso mixto de ambos (Rutter y Tuma, 1988; Jensen, Watanabe, Richters, Roper, Hibbs, Salzberg y Liu, 1996).

Respecto al punto de corte, los autores del CBCL aconsejan un punto de corte de la puntuación total T del CBCL entre 60 y 63. Nuestro trabajo sugiere que el punto de corte óptimo en población de procedencia mixta debe establecerse por debajo de estos valores. En su uso más habitual, como instrumento de cribado, un punto de corte de 53-54 proporcionaría valores aceptables de sensibilidad y especificidad. En estudios de prevalencia, en los que se requiere detectar un elevado porcentaje de sujetos patológicos, el punto de corte debería fijarse entre 46 y 50. El elevado número de sujetos sanos clasificados erróneamente como patológicos sería detectado en una segunda fase con pruebas de evaluación específicas de la patología sugerida por el CBCL (Fletcher, 1988).

Hemos abordado pues un punto crucial en psicopatología: el criterio o patrón de referencia o *gold standard*, y hemos comprobado que un instrumento de cribado transculturalmente aceptado y validado (Bird, Gould, Rubio-Stipec y Staghezza, 1991; Carter, Grogorenko y Pauls, 1995; Hellinckx, Grietens y Verhulst, 1994; Schneider, Walter y Remschmidt, 1991; Verhulst, Achenbach, Ferdinand y Kasius, 1993; Xin, Chen, Tang, Lin y McConville, 1992; Zilber, Auerbach y Lerner, 1994) es más o menos eficaz en función de la calidad del criterio. También hemos puesto de manifiesto que el análisis de curvas ROC es la metodología más adecuada para evaluar la precisión diagnóstica de un instrumento de cribado.

REFERENCIAS

- Achenbach, T.M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

- Bird, H.R., Gould, M.S., Rubio-Spítec, M. & Staghezza, B. (1991). Screening for childhood psychopathology in the community using the Child Behavior Checklist. *Journal of the American Academy of Child and Adolescent Psychiatry*, 30, 1, 116-123.
- Carter, A.S., Grigorenko, E.L. & Pauls, D.L. (1995). A Russian adaptation of the Child Behavior Checklist: Psychometric properties and associations with child and maternal affective symptomatology and family functioning. *Journal of Abnormal Child Psychology*, 23, 6, 661-684.
- Ezpeleta, L., de la Osa, N., Doménech, J.M., Navarro, J.B. y Losilla, J.M. (1995). La *Diagnostic Interview for Children and Adolescents - DICA-R*: Acuerdo diagnóstico entre niños/adolescentes y sus padres. *Revista de Psiquiatría de la Facultad de Medicina de Barcelona*, 22, 153-163.
- Ezpeleta, L., de la Osa, N., Júdez, J., Doménech, J.M., Navarro, J.B. & Losilla, J.M. (in press). Diagnostic agreement between clinician and the Diagnostic Interview for Children and Adolescents - DICA-R in a Spanish outpatient sample. *Journal of Child Psychology and Psychiatry*.
- Ezpeleta, L. (1994). *Checklist*. Puede obtenerse en el Departamento de Psicología de la Salud, Universitat Autònoma de Barcelona.
- Fletcher, R.H. (1988). *Clinical epidemiology*. Baltimore: Williams & Wilkins.
- Fombonne, E. (1991). The use of questionnaires in child psychiatry research: Measuring their performance and choosing an optimal cut-off. *Journal of Child Psychology and Psychiatry*, 32, 677-693.
- Godoy, A. (1996). *Toma de decisiones y juicio clínico*. Madrid: Pirámide.
- Green, D.M. & Swets, J.A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley and Sons, Inc.
- Hanley, J.A. & McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29-36.
- Hanley, J.A. & McNeil, B.J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148, 839-843.
- Hellinckx, W., Grietens, H. & Verhulst, F. (1994). Competence and behavioral problems in 6 to 12 year old children in Flanders (Belgium) and Holland: A cross-national comparison. *Journal of Emotional and Behavioral Disorders*, 2, 3, 130-142.
- Hollingshead, A.B. (1975). *Four factor index of social status*. Manuscrito no publicado. Department of Sociology, Yale University.
- Jensen, P.S., Watanabe, H.K., Richters, J.E., Roper, M., Hibbs, E.D., Salzberg, A.D. & Liu, S. (1996). Scales, diagnoses, and child psychopathology: II. Comparing the CBCL and the DISC against external validators. *Journal of Abnormal Child Psychology*, 24, 151-168.
- Lane, J.W., Pollard, A. & Cox, G.I. (1990). Validity study of the anxiety symptoms interview. *Journal of Clinical Psychology*, 46, 52-57.
- de la Osa, N., Ezpeleta, L., Doménech, J.M., Navarro, J.B. y Losilla, J.M. (1996a). Fiabilidad entre entrevistadores de la Entrevista Diagnóstica Estructurada para Niños y Adolescentes (DICA-R). *Psicothema*, 8, 359-368.
- de la Osa, N., Ezpeleta, L., Doménech, J.M., Navarro, J.B. y Losilla, J.M. (1996b). *Validez convergente y discriminante de la Entrevista Diagnóstica Estructurada DICA-R*. Manuscrito enviado para publicación.
- Reich, W., Shayka, J.J. & Taibleson, Ch. (1991). *Diagnostic Interview Schedule for Children and Adolescent - DICA-R*. Unpublished manuscript, Washington University, Division of Child Psychiatry.
- Robins, L. (1985). Epidemiology: Reflections on testing the validity of psychiatric interviews. *Archives of General Psychiatry*, 42, 918-924.
- Rutter, M. & Tuma, A.H. (1988). Diagnosis and classification: some outstanding issues. In M. Rutter, A.H. Tuma & I.S. Lann (Eds.), *Assessment and diagnosis in child psychopathology* (pp. 437-452). London: David Fulton Publishers.
- Schneider, K., Walter, R., & Remschmidt, H. (1991). Validation of a German version of the Child Behavior Checklist (CBCL). *Zeitschrift für Klinische Psychologie, Forschung und Praxis*, 20, 1 52-64.
- Simpson, A.J. & Fitter, M.J. (1973). What is the best index of detectability? *Psychological bulletin*, 80, 481-488.
- Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285-1293.
- Turner, D.A. (1978). An intuitive approach to receiver operating characteristic curve analysis. *The Journal of Nuclear Medicine*, 19, 213-220.
- Vaillant, G.E. & Schnurr, P. (1988). What is a case? *Archives of General Psychiatry*, 45, 313-319.
- Verhulst, F.C. & Koot, H.M. (1992). *Child psychiatric epidemiology* (Developmental clinical psychology and psychiatry series) (Vol. 23). Newbury Park, California: SAGE Publications.
- Verhulst, F.C., Achenbach, T.M., Ferdinand, R.F. & Kasius, M.C. (1993). Epidemiological comparisons of American and Dutch adolescents' self-reports. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32, 1135-1144.
- Weinstein, M.C. & Fineberg, H.V. (1980). *Clinical Decision Analysis*. Philadelphia: W.B. Saunders Company.

- Xin, R., Chen, S.K., Tang, H.Q., Lin, X.F. & McConville, B.J. (1992). Behavioral problems among preschool age children in Shanghai: Analysis of 3.000 cases. *Canadian Journal of Psychiatry*, 37, 250-258.
- Zilber, N., Auerbach, J. & Lerner, Y. (1994). Israeli norms for the Achenbach Child Behavior Checklist: Comparison of clinically-referred and non-referred children. *Israel Journal of Psychiatry and Related Sciences*, 31, 1, 5-12.

