

## ARTÍCULO DE INVESTIGACIÓN

## Pruebas de tamizaje de neurodesarrollo global para niños menores de 5 años de edad validadas en Estados Unidos y Latinoamérica: revisión sistemática y análisis comparativo

### *Global neurodevelopmental screening tests for children under 5 years of age in the United States and Latin America: systematic review and comparative analysis*

Beatriz Romo-Pardo, Silvia Liendo-Vallejos, Guillermo Vargas-López, Antonio Rizzoli-Córdoba, Guillermo Buenrostro-Márquez

#### RESUMEN

**Introducción.** La Academia Americana de Pediatría recomienda aplicar, sistemáticamente, pruebas de tamizaje del neurodesarrollo para el oportuno diagnóstico y tratamiento. Para lograr referir apropiadamente a los pacientes, es importante tener información bien fundamentada sobre estas herramientas.

**Métodos.** Se realizó una revisión sistemática de estudios de validación de pruebas de tamizaje de neurodesarrollo para menores de 5 años en Estados Unidos y Latinoamérica, de 1980 a 2012.

**Resultados.** Se encontraron 13 pruebas con diferentes características de evaluación y calificación. Obtuvieron la mejor sensibilidad y especificidad *Battelle Developmental Inventory Screening (2nd Edition)* (0.93 y 0.88) y PRUNAPE con VPP 0.94 y VPN 0.97.

**Conclusiones.** De 1980 a 2012, se encontraron 13 pruebas de tamizaje del neurodesarrollo para menores de 5 años en Estados Unidos y Latinoamérica. La *Battelle Developmental Inventory Screening* obtuvo la mayor validez de criterio, y la PRUNAPE fue la más predictiva. No se encontraron estudios de validación en México. Por lo tanto, se consideró de vital importancia contar con una herramienta validada en nuestro país.

**Palabras clave:** neurodesarrollo, pruebas de tamizaje, validación.

#### INTRODUCCIÓN

La detección temprana de los problemas en el neurodesarrollo es de suma importancia para el bienestar de los niños y sus familias, ya que permite acceder a un diagnóstico y tratamiento oportunos.<sup>1</sup> En los países en vías de desa-

#### ABSTRACT

**Background.** The American Academy of Pediatrics recommends the application of neurodevelopmental screening tests for early intervention of neurodevelopmental disorders. In order to refer these tests appropriately, it is important to have well-founded information in regard to these tools.

**Methods.** A systematic literature search targeted on validation studies of neurodevelopmental screening tests in children <5 years of age in the U.S. and Latin America from 1980 to 2012 was conducted.

**Results.** We found 19 validation studies of 13 screening tests. *Battelle Developmental Screening Inventory (2nd edition)* reported the best sensitivity and specificity (0.93/0.88) and PRUNAPE with predictive positive and negative values (0.94/0.97).

**Conclusions.** From 1980-2012 we found 13 neurodevelopmental screening tests in the U.S. and Latin America for children <5 years of age. The best criterion and predictive validity was for *Battelle Developmental Inventory Screening* and PRUNAPE, respectively. No validation studies were found in Mexico; therefore, we consider it important to have a validated tool in our country.

**Key words:** neurodevelopment, screening tests, validation.

rollo, un gran número de niños menores de 5 años están expuestos a múltiples factores de riesgo, como la pobreza, la desnutrición, los problemas de salud y un ambiente con pobre estimulación, lo cual afecta su desarrollo cognitivo, motor y socio emocional.<sup>2</sup> Se ha observado que los niños que reciben una intervención temprana presentan, a largo plazo, una mejoría en el coeficiente intelectual, mejor desempeño escolar, menor índice de criminalidad y, en la edad adulta, una mayor posibilidad de obtener empleo e ingresos más elevados en comparación con aquellos que no la recibieron.<sup>3</sup>

Para identificar las alteraciones en el neurodesarrollo, la Academia Americana de Pediatría sugiere una vigilancia y

Dirección de Investigación,  
Hospital Infantil de México Federico Gómez  
México D.F., México

Fecha de recepción: 03-10-12  
Fecha de aceptación: 26-10-12

monitoreo continuo del desarrollo, tomando en cuenta los factores de riesgo tanto biológicos como ambientales, así como las preocupaciones de los padres sobre el desarrollo de sus hijos en cada visita de seguimiento. Otra recomendación es la aplicación sistemática de pruebas de tamizaje en momentos claves del desarrollo, esto es, a los 9, 18 y 30 meses de edad.<sup>1</sup> Diversos estudios han demostrado que el juicio clínico del pediatra no es suficiente para identificar retrasos en el neurodesarrollo. De esto, la importancia de emplear herramientas de tamizaje estandarizadas para detectar estos pacientes.<sup>4,5</sup>

Una prueba de tamizaje identifica a los individuos presuntamente enfermos en una población aparentemente sana; establece el riesgo o la sospecha de algún problema del desarrollo, pero no define un diagnóstico. Debe ser fácil y rápida de aplicar, económicamente viable, confiable y válida (especificidad y sensibilidad >0.70).<sup>4</sup> La utilidad de una prueba debe ir precedida de un proceso de validación concurrente, donde se busca conocer la medida en la que coinciden los resultados de dicha prueba con las evaluaciones diagnósticas habitualmente utilizadas (estándar de oro).<sup>6,7</sup>

Contar con una evidencia sólida sobre las pruebas de tamizaje del neurodesarrollo es facilitar la comprensión de sus ventajas y desventajas como una herramienta de evaluación. Con el objetivo de tener un mejor conocimiento acerca de la efectividad y exactitud de las pruebas de tamizaje en niños, así como proporcionar información útil sobre estas herramientas basada en evidencias bien fundamentadas, se realizó una revisión sistemática de la literatura sobre la validación de las pruebas de tamizaje diseñadas en Estados Unidos y Latinoamérica para la detección de problemas de desarrollo global en menores de 5 años de edad.

## MÉTODOS

Para los criterios de inclusión y exclusión, se realizó una búsqueda y revisión sistematizada en la literatura de los estudios de validación de las pruebas de tamizaje de neurodesarrollo global para niños menores de 5 años de edad en Estados Unidos y Latinoamérica, desde 1980 y hasta el 14 de febrero de 2012, en inglés y español. Quedaron excluidos los trabajos de revisión, editoriales y comunicaciones a congresos.

Los estudios fueron identificados por una búsqueda en las bases de datos electrónicas y la bibliografía de los artículos encontrados. Esta búsqueda se realizó en MEDLINE/PubMed, LILACS y ARTEMISA. La búsqueda se limitó a estudios en inglés o español, que fueron realizados en humanos, en menores de 5 años de edad, en Estados Unidos y Latinoamérica, en la categoría de estudios de validación.

Se combinaron los siguientes sets de términos de búsqueda:

1. Para la enfermedad: retraso en el desarrollo
2. Para el tipo de estudios: tamizaje
3. Para el desenlace: propiedades psicométricas, sensibilidad, especificidad
4. Para el personal que aplica el tamizaje: atención primaria, médico general

A continuación se muestran ejemplos de las estrategias de búsqueda de las diferentes combinaciones de términos en PubMed/MEDLINE:

- Estrategia 1: 184 artículos  
("diagnosis"[Subheading] OR "diagnosis"[All Fields] OR "screening"[All Fields] OR "mass screening"[MeSH Terms] OR ("mass"[All Fields] AND "screening"[All Fields]) OR "mass screening"[All Fields] OR "screening"[All Fields] AND ("sensitivity and specificity"[MeSH Terms] OR ("sensitivity"[All Fields] AND "specificity"[All Fields]) OR "sensitivity and specificity"[All Fields] OR "sensitivity"[All Fields] AND (developmental[All Fields] AND delay[All Fields]))
- Estrategia 2: 154 artículos; 143 artículos + 11 incluidos en la estrategia 1  
("diagnosis"[Subheading] OR "diagnosis"[All Fields] OR "screening"[All Fields] OR "mass screening"[MeSH Terms] OR ("mass"[All Fields] AND "screening"[All Fields]) OR "mass screening"[All Fields] OR "screening"[All Fields] AND ("americas"[MeSH Terms] OR "americas"[All Fields] OR "america"[All Fields]) AND (developmental[All Fields] AND delay[All Fields]) antiguo
- Estrategia 3: 74 artículos; 53 artículos + 21 incluidos en las estrategias 1 y 2  
("diagnosis"[Subheading] OR "diagnosis"[All Fields] OR "screening"[All Fields] OR "mass screening"[MeSH Terms] OR ("mass"[All Fields] AND "screening"[All Fields]) OR "mass screening"[All Fields] OR "screening"[All Fields] AND ("primary health care"[MeSH

Terms] OR ("primary"[All Fields] AND "health"[All Fields] AND "care"[All Fields]) OR "primary health care"[All Fields] OR ("primary"[All Fields] AND "care"[All Fields]) OR "primary care"[All Fields] AND (developmental[All Fields] AND delay[All Fields])

- Estrategia 4: 47 artículos; 36 artículos + 11 incluidos en las estrategias 1, 2 y 3  
("diagnosis"[Subheading] OR "diagnosis"[All Fields] OR "screening"[All Fields] OR "mass screening"[MeSH Terms] OR ("mass"[All Fields] AND "screening"[All Fields]) OR "mass screening"[All Fields] OR "screening"[All Fields] AND ("psychometrics"[MeSH Terms] OR "psychometrics"[All Fields] OR "psychometric"[All Fields]) AND (developmental[All Fields] AND delay[All Fields]) AND ((English[lang] OR Spanish[lang]) AND ("infant"[MeSH Terms] OR "child, preschool"[MeSH Terms]))

- Estrategia 5: 12 artículos; 9 artículos + 3 incluidos en la estrategia 3  
("diagnosis"[Subheading] OR "diagnosis"[All Fields] OR "screening"[All Fields] OR "mass screening"[MeSH Terms] OR ("mass"[All Fields] AND "screening"[All Fields]) OR "mass screening"[All Fields] OR "screening"[All Fields] AND ("general practitioners"[MeSH Terms] OR ("general"[All Fields] AND "practitioners"[All Fields]) OR "general practitioners"[All Fields] OR ("general"[All Fields] AND "physician"[All Fields]) OR "general physician"[All Fields]) AND (developmental[All Fields] AND delay[All Fields])).

La evaluación de la elección de los artículos incluidos fue realizada de forma independiente y abierta por dos revisores. Los desacuerdos entre ellos fueron resueltos por consenso. Se desarrolló una hoja de captura de datos basados en la plantilla de extracción de datos del grupo de usuarios y revisores de comunicación de Cochrane. Se definió como prueba de tamizaje de neurodesarrollo global a aquellas que evaluaron múltiples dominios del desarrollo (por ejemplo motor, cognitivo, adaptativo, comunicación, entre otros). Se excluyeron aquellas pruebas que se enfocaron en una única área del desarrollo, aquellas encaminadas al diagnóstico de una enfermedad o que evaluaran áreas académicas. La información obtenida de cada prueba fue

la siguiente: 1) características de la prueba (nombre, país de origen, autores, modalidad de evaluación, dominios del desarrollo, rango de edad en que se aplica, sistema de calificación, criterios de normalidad y anormalidad y tiempo de evaluación); 2) prueba diagnóstica utilizada para su validación; y 3) resultados de validación.

## RESULTADOS

Con respecto a los resultados de la revisión sistemática en las diferentes estrategias de búsqueda se encontraron 454 artículos en total, de los cuales se incluyeron únicamente 19 artículos que describen 13 pruebas de tamizaje del neurodesarrollo (Figura 1).

Los 19 artículos seleccionados fueron diversos estudios (tanto de validación como revisiones sistemáticas) de las pruebas de tamizaje del neurodesarrollo publicados en el continente americano. Se utilizaron aquellos en idioma inglés o español. De las 13 pruebas encontradas, las que tuvieron un mayor número de publicaciones fueron *Ages and Stages Questionnaires* (5 artículos) y CAT/CLAMS (3 artículos). El estudio más antiguo incluido fue de 1986 de la prueba de tamizaje CAT/CLAMS. Un total de 9,217 niños menores de 8 años de edad fueron englobados en los diversos estudios de validación de las pruebas realizados en Argentina, Canadá, Chile, Costa Rica, Cuba y Estados Unidos. Es interesante mencionar que en la revisión sistemática se encontró un único artículo en donde se evaluaron las características psicométricas de una prueba de tamizaje en población mexicana; se trata de un estudio de estandarización de la prueba de Denver I en 288 niños de 2 a 54 semanas de vida, donde se utilizó también la escala motora de la prueba de Bayley. La escala de Denver I falló en identificar como sospechosos a 16 de los 17 niños identificados por la prueba de Bayley. El estudio se consideró como no satisfactorio y en la práctica no se utilizan los resultados como estándares mexicanos. Se decidió no incluir este artículo en esta revisión sistemática, ya que actualmente se utiliza la segunda versión de Denver.

Por su forma de administración, las pruebas de tamizaje seleccionadas pueden categorizarse en dos grandes grupos: las de observación o evaluación directa (realizada por el médico al niño) y los cuestionarios para padres (que puede aplicar cualquier personal de salud). Hay pruebas que utilizan ambos recursos, en las que se realizan preguntas a los padres y además se observa al niño. Los ítems que

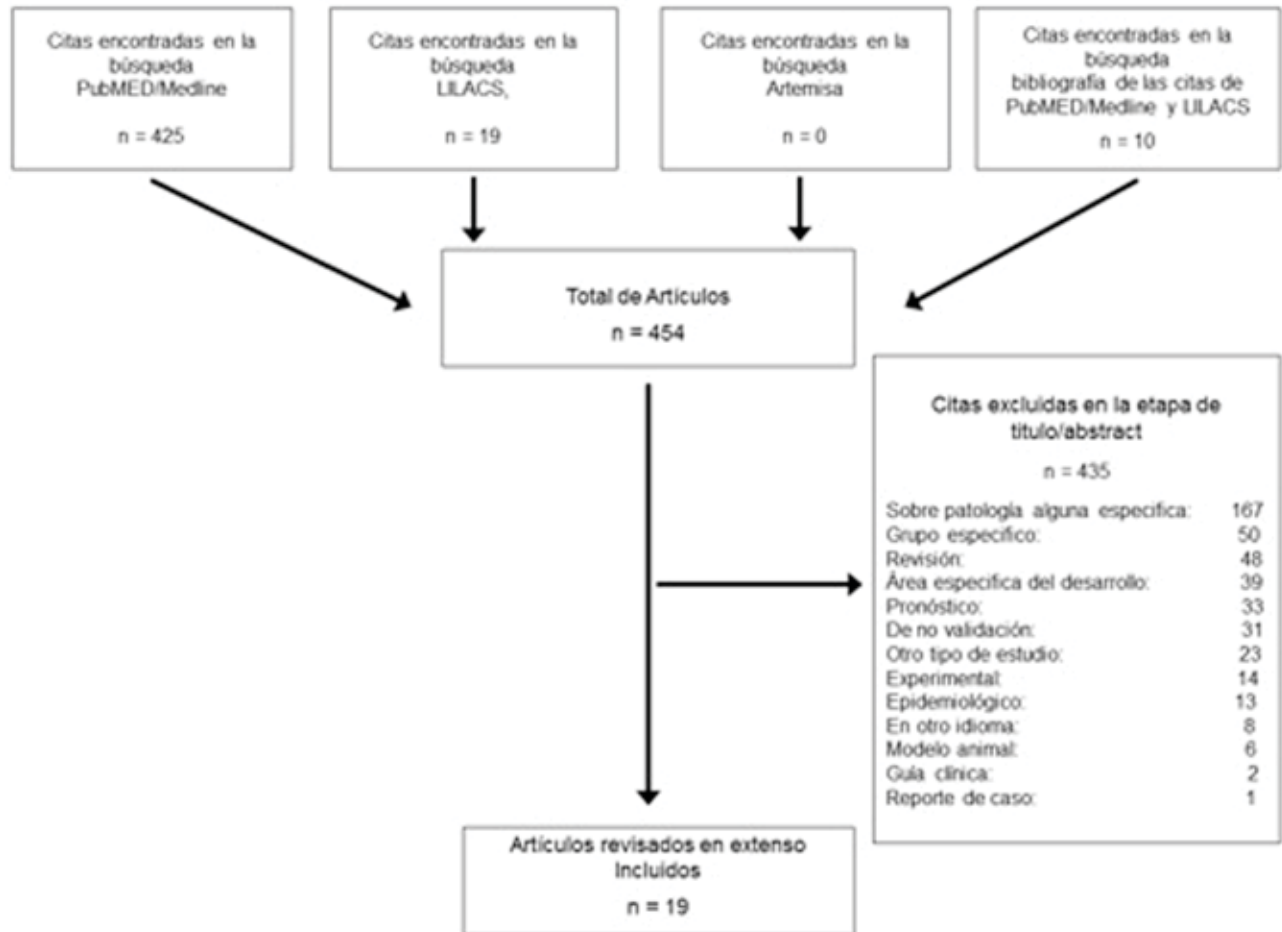


Figura 1. Diagrama de flujo de los resultados de la revisión sistemática.

evalúa cada prueba están distribuidos en diferentes áreas de desarrollo. Si bien existe bastante homogeneidad en la agrupación de los hitos motores o de lenguaje, en los dominios de conducta adaptativa y social los ítems están distribuidos de manera diferente en las distintas pruebas. Se describieron las diferentes pruebas de tamizaje encontradas, con la mención de sus autores, país de origen, modalidad de administración, dominios del desarrollo que evalúa, tiempo de administración e idiomas disponibles (Cuadro 1).

Los sistemas de calificación varían en las diferentes pruebas. El puntaje total se obtiene generalmente a partir de los puntajes individuales de los ítems. Estos, a su vez, se desprenden de las respuestas de los padres a las preguntas del cuestionario (por ejemplo, el *Ages & Stages Questionnaires*) o de la puntuación otorgada por el

médico al desempeño del niño (el *Battelle Developmental Inventory*). En general, las pruebas de tamizaje califican al niño como normal o como sospechoso o con riesgo para problemas desarrollo. En el Cuadro 2 se observa el método de calificación y los criterios de normalidad y anormalidad para cada prueba.

La eficacia de las medidas de detección se reportó en los estudios de validación. El concepto de validez se refiere a qué tan bien una herramienta mide lo que pretende medir. Para la validación concurrente de las pruebas de tamizaje se han utilizado diferentes instrumentos de diagnóstico, como son el Inventario de desarrollo de Battelle (*Battelle Developmental Inventory*), la Escala de desarrollo infantil de Bayley (*Bayley Scales of Infant Development*), la Escala preescolar y primaria de Wechsler (*Wechsler Preschool and Primary Scale of Intelligence*), entre otras.

**Cuadro 1.** Descripción general de las pruebas de tamizaje según la modalidad de administración, rango de edad, tiempo de administración e idiomas disponibles (Continúa en la siguiente página)

<i>Nombre de la prueba País de origen Autores</i>	<i>Modalidad de evaluación</i>	<i>Dominios del desarrollo evaluados</i>	<i>Rango de edad en meses</i>	<i>Tiempo</i>	<i>Idioma</i>
Ages & Stages Questionnaires 3th ed. (EUA) Squires, et al, 1995 <sup>8</sup> Squires, et al, 2009 <sup>9</sup>	Cuestionario para padres	Comunicación, motricidad gruesa, motricidad fina, resolución de problemas, lenguaje, personal social	4-60	10-15 minutos	Inglés, espa- ñol, francés y coreano
Battelle Developmental Inven- tory (BDI) Screening Test 2nd ed. (EUA) Newborg J, 2005 <sup>12</sup>	Observación directa del niño y preguntas a los padres	Personal-social, adaptati- vo, motor, comunicación y cognitivo	0-95	10-30 minutos	Inglés y español
Bayley Scales of Infant and Toddler Development (BSID) Neurodevelopment Screening Test 3 <sup>rd</sup> ed. (EUA) Aylward G, 2010 <sup>13</sup>	Observación directa del niño	Cognitivo, lenguaje y motor	1-42	15-25 minutos	Inglés
Brigance Early Childhood Screen (EUA) Glascoe F, 2002 <sup>14</sup>	Observación directa del niño y cuestiona- rios para padres	Cognitivo, lenguaje, mo- tor, adaptativo y socioe- mocional	0-35 36-60	10-15 minutos	Inglés y español
CAT/CLAMS Clinical Adapta- tive Test/Clinical Linguistic and Auditory Milestone Scale (EUA) Capute, et al, 1986 <sup>15</sup>	Evaluación directa del niño	Lenguaje, resolución de problemas y motor	1-36	10-15 minutos	Inglés y español
Child Development Inventory (EUA) Doig, et al, 1999 <sup>16</sup>	Cuestionario para padres	Social, lenguaje, motor, adaptativo, habilidades de lectura y aritmética	15-72	30-50 minutos	Inglés
Denver Development Scree- ning Test (EUA) Glascoe, et al, 1992 <sup>17</sup>	Evaluación directa del niño y preguntas a los padres	Motricidad gruesa, lenguaje, motricidad fina-adaptativa, personal- social	0-72	10-20 minutos	Inglés y español
Escala de Evaluación del Desarrollo Psicomotor (EEDP) (Chile) Schapira, 2007 <sup>18</sup> Bedregal, 2008 <sup>19</sup> Vericat Ay Orden, 2010 <sup>20</sup>	Evaluación directa del niño	Social, lenguaje, coordi- nación y motricidad	0-24	20 minutos	Español
Escala de Desarrollo Integral del Niño (EDIN) (Costa Rica) Schapira, 2007 <sup>18</sup> Vericat Orden, 2010 <sup>20</sup>	Evaluación directa del niño	Motricidad fina, motricidad gruesa, reflejos, socio- emocional y cognitiva	0-72	NR	Español
Neurodesarrollo Pediátrico (NPED) (Cuba) Guadarrama-Celaya, et al, 2011 <sup>21</sup>	Evaluación directa del niño	Lenguaje/comunicación, psicomotor y maduración sensorial (audición/visión)	1-60	15 minutos	Español

**Cuadro 1.** Descripción general de las pruebas de tamizaje según la modalidad de administración, rango de edad, tiempo de administración e idiomas disponibles (Continuación)

Nombre de la prueba País de origen Autores	Modalidad de evaluación	Dominios del desarrollo evaluados	Rango de edad en meses	Tiempo	Idioma
PEDS Parents Evaluation of Developmental Status (EUA) Glascoe, 1998	Cuestionario para padres que pregunta sobre preocupaciones	Desarrollo global/cognitiva, lenguaje expresivo, lenguaje receptivo, conducta, socio-emocional, escolaridad, autoayuda, motor fino y grueso y otros (preocupaciones sensoriales/médicas)	0-96	2-5 minutos	Inglés y español
PRUNAPE (Argentina) Pascucci, et al, 2002 <sup>22</sup> ; 2006 <sup>23</sup>	Evaluación directa del niño y algunas preguntas a los padres	Motricidad fina, motricidad gruesa, personal social y lenguaje	0-60	10-15 minutos	Español
Test de desarrollo Psicomotor TEPSI (Chile) Haeussler, Marchant, 1980	Evaluación directa del niño	Coordinación, motricidad y lenguaje	24-60	15-20 minutos	Español

Se examinaron tanto la validez de criterio como la validez predictiva.

#### Validez de criterio

La validez de criterio es un tipo de validez concurrente que establece la validez de un instrumento de medición al compararla con algún criterio externo. La sensibilidad muestra que tan bien una prueba identifica correctamente a los niños con retraso, mientras que la especificidad indica el grado por el cual una prueba detecta a aquellos sin retraso. Algunas pruebas mostraron ser una herramienta de selección deficiente para reconocer a los niños con retraso en el neurodesarrollo, ya que mostraron una sensibilidad de 0.50, como la prueba de *Child Development Inventory* (CDI); o ser incapaz de diferenciar a aquellos con un neurodesarrollo normal en comparación con uno anormal—con una especificidad de 0.43-0.80—, como la prueba de Denver. Por el contrario, otros estudios demostraron ser un instrumento útil en la evaluación del neurodesarrollo tras mantener una alta sensibilidad y especificidad como la *Battelle Developmental Inventory Screening (2nd Edition)*, con una sensibilidad de 0.93 y especificidad de 0.88.

#### Validez predictiva

La validez predictiva es un tipo de validez concurrente que se refiere a la capacidad que tiene una prueba para

predecir o correlacionar con otra del mismo constructo. Se encontraron 8 estudios con validez predictiva. El peor valor predictivo fue para *Ages & Stages Questionnaires* con 0.34 valor predictivo positivo (VPP) y 0.71 de valor predictivo negativo (VPN); mientras que la prueba con mejor validez predictiva fue PRUNAPE con VPP de 0.94 y VPN de 0.97 (Cuadro 3).

Cada una de las pruebas de tamizaje del neurodesarrollo seleccionadas para esta revisión tiene ventajas y desventajas en cuanto a su método y tiempo de aplicación, materiales y características del estudio de validación. Se han resumido las fortalezas y debilidades de cada una de las pruebas, así como algunas observaciones sobre el proceso de validación (Cuadro 4).

## DISCUSIÓN

Las recomendaciones sobre la aplicación sistemática de pruebas de tamizaje del neurodesarrollo sugeridas por la Academia Americana de Pediatría han propiciado un aumento en la demanda de pruebas que contengan evidencias, confiabilidad, utilidad, validez, especificidad y sensibilidad sólidas. Estar al tanto de la evidencia disponible sobre estas herramientas conlleva a una toma de decisiones por parte de los médicos provista de fundamento. Es útil conocer como una investigación sobre un instrumento se llevó

**Cuadro 2.** Método de calificación y criterios de anormalidad y normalidad de las pruebas de tamizaje incluidas en esta revisión sistemática (Continúa en la siguiente página)

<i>Prueba de tamizaje</i>	<i>Sistema de calificación</i>	<i>Criterio de normalidad y anormalidad</i>
Ages & Stages Questionnaires (EUA)	Preguntas para padres con opción de respuesta: sí, algunas veces o aún no Dando un puntaje de 0 para no, 5 para algunas veces y 10 para sí Además se evalúan las preocupaciones de los padres con respuestas de sí y no	Se divide en 3 categorías (arriba del punto de corte, cerca del punto de corte y debajo de corte)  El punto de corte esta entre 1.5-2 DE Con esto determina la posibilidad de una valoración subsecuente
Battelle Developmental Inventory Screening 2nd ed (EUA)	Son 100 ítems (20 de cada dominio del desarrollo: motor, comunicación, cognición, conducta adaptativa y social). El puntaje de 0, 1 y 2 en cada uno refleja el nivel de adquisición de las habilidades	ANORMALIDAD:  < -1 DE y > -1.5 DE: límite. Referir al niño para mayor evaluación con un BDI-2  < -1.5 DE y > -2.0 DE: clara indicación de referencia. Administrar BDI-2 para determinar áreas específicas de déficit  < -2.0 DE: clara indicación de problemas severos de desarrollo. Administrar el BDI-2 y determinar la naturaleza y extensión del déficit
Bayley Scale of Infant and Toddler Development Screening Test (EUA)	Se califican 4 áreas (función neurológica básica/intacta, funciones receptivas, funciones expresivas y proceso cognitivos) con opciones de calificación 1 (óptimo) y 2 (no óptimo)	Dos puntos de corte que divide en tres categorías: riesgo leve, moderado y severo  No da un puntaje total de desarrollo, únicamente para cada una de 4 áreas evaluadas
Brigance Screens-II (EUA)	Los resultados son basados en el criterio del examinador	Sin criterios establecidos para clasificar
CAT/CLAMS (EUA)	Dos dominios: comunicación y resolución de problemas Ítems organizados por grupo de edad con puntuación (0.3-1.5) Puntuación total por dominio y puntuación general de la prueba	Establece edad basal, edad techo, edad equivalente y cociente de desarrollo  Punto de corte, cociente de desarrollo $\leq 70$ : retraso en el desarrollo
Child Development Inventory (EUA)	Ocho dominios calificados por los padres con opción de respuesta sí o no	Clasifica en 3 grupos: Límites normales, límite $< 1.5$ DE y con retraso $< 2$ DE
DENVER-II (EUA)	Ítems administrados al niño o por información obtenida por los padres de acuerdo a la línea de edad Se clasifica cada ítem: realizado con éxito, falla o rechaza hacerlo	Normal: habilidades para su edad (1 falla por área) Sospecha: falla para realizar las habilidades que realizan 75-95% de los niños de su edad ( $> 2$ fallas en más de 2 áreas) Retraso: falla para realizar las actividades que realizan más del 90% de los niños de su edad
EEDP Escala de Evaluación del Desarrollo Psicomotor (Chile)	Ítems administrados al niño. Puntuación 0 si falla, 1 o 2 si pasa	Suma y punto de corte. Se establecen las categorías: normal, riesgo, retraso
NPED Neurodesarrollo Pediátrico (Cuba)	Instrumento computarizado de evaluación que pueden contestar enfermeras	Se reporta como: normal, fracaso global o por áreas
Parents Evaluation of Developmental Status (EUA)	El padre responde: sí, no y un poco. Contesta preguntas acerca de preocupaciones sobre el desarrollo global, el lenguaje expresivo y receptivo, la motricidad fina y gruesa, la conducta y el interés social	Clasifica en 3 categorías de riesgo de retraso en el desarrollo: leve, moderado y severo

**Cuadro 2.** Método de calificación y criterios de anormalidad y normalidad de las pruebas de tamizaje incluidas en esta revisión sistemática (Continuación)

<i>Prueba de tamizaje</i>	<i>Sistema de calificación</i>	<i>Criterio de normalidad y anormalidad</i>
PRUNAPE (Argentina)	Se administran pautas al niño y se realizan preguntas a los padres de acuerdo a la línea de edad. Cada pauta se califica como: pasa (si aprueba el criterio) o falla (si no aprueba) Las pautas tipo A son las que están completamente a la izquierda de la línea de edad. Las pautas tipo B son atravesadas por la línea de edad en el percentil 75-90%	Pasa la prueba si cumple todas las pautas tipo A y falla en hasta una pauta tipo B El niño se califica como sospechoso o de riesgo si fracasa en una pauta tipo A o en 2 pautas tipo B
TEPSI Test de evaluación del desarrollo psicomotriz	Evaluación del niño con ítems que se califican con puntuación 0, 1 y 2	Suma y punto de corte. Se establecen las categorías: normal, riesgo, retraso

**Cuadro 3.** Comparación de pruebas de tamizaje diseñadas y validadas según los resultados de la búsqueda sistemática de la literatura (Continúa en la siguiente página)

<i>Prueba validada</i>	<i>Prueba diagnóstica utilizada para comparación</i>	<i>Resultados de validación</i>
Ages & Stages Questionnaires (EUA) Squires & Bricker, 1997	BSID SBIS MSCA	S=0.70-0.90 E=0.76-0.91 VPP=0.45
Ages & Stages Questionnaires (EUA) Limbo MM, Joyce DP; 2011	BSID WPPSI-III VABS PLS-IV	S=0.82 E=0.78 VPP=0.30 ( $\geq 1$ dominio) y 0.48 ( $\geq 2$ dominios) VPN=0.97 ( $\geq 1$ dominio) y 0.94 ( $\geq 2$ dominios)
Ages & Stages Questionnaires (Canada) Rydz, 2006	BDI-2	S=0.67 E=0.39 VPP=0.34 VPN=0.71
BDI Battelle Developmental Inventory Screening 2 <sup>nd</sup> edition (EUA) Newborg J, 2005	BDI-2	S=0.72-0.93 E=0.79-0.88
BSID Bayley Infant Neurodevelopmental Screen (EUA) Aylward GP, 2005	BSID	S=0.61-0.80 E=0.81-0.90
Brigance Screens-II Glascoe FP, 2002	Panel multidisciplinario de especialistas: pediatras, enfermeras, maestras y psicólogos de desarrollo Reporte de padres	S=0.76-0.77 E=0.85-0.86
CAT/CLAMS Capute, 1986	BSID	S=0.21- 0.66 E=0.79-0.95 VPP=0.80 VPN=0.65
CDI Child Development Inventory (EUA) Doig, et al, 1999	CAT/CLAMS y BSID	S=0.70-0.80 E=0.70-0.80
CDI Child Development Inventory Rydz, et al, 2006	BDI	S=0.50 E=0.86
DENVER-II (EUA) Glascoe, et al, 1992	BSID KABC SBIS VABS	S=0.56-0.83 E=0.43- 0.80 VPP=0.37



**Cuadro 3.** Comparación de pruebas de tamizaje diseñadas y validadas según los resultados de la búsqueda sistemática de la literatura (Continuación)

<i>Prueba validada</i>	<i>Prueba diagnóstica utilizada para comparación</i>	<i>Resultados de validación</i>
EDIN Escala de Evaluación Integral del Niño (Costa Rica) Schapira, 2007 Vericat, 2010	No hay datos disponibles sobre validación	
EEDP Escala de Evaluación del Desarrollo Psicomotor (Chile) Schapira, 2007 Bedregal, 2008 Vericat A, 2010	No se encontraron los datos de la validación	Baja S y E en menores de 4 meses VPP en mayores de 4 meses: 97-100%
NPED Neurodesarrollo Pediátrico (Cuba) Guadarrama-Celaya, et al, 2011	No se encontraron datos detallados de la validación	S=0.95 E=0.86
PEDS Parents Evaluation of Developmental Status (EUA) Glascoe, et al, 2003 Glascoe, 2001	El desarrollo de los niños se midió empleando un conjunto de pruebas estándar por parte de licenciados en psicología de forma ciega, bien con respecto a las preocupaciones de los padres o a su relevancia	S=0.74-0.79 E=0.70-0.80
PEDS Parents Evaluation of Development Status (EUA) Limbos MM, 2011	BSID WPPSI-III VABS PLS-IV	S=0.74 E=0.64 VPP=0.19 ( $\geq 1$ preocupación) y 0.30 ( $\geq 2$ preocupaciones) VPN=0.96 ( $\geq 1$ preocupación) y 0.93 ( $\geq 2$ preocupaciones)
PRUNAPE (Argentina) Pascucci MC, 2002 y 2004	BSID WISC Terman VABS Prueba de Gardner EN Evaluación clínica por psiquiatría EA Audiometría tonal PEATC	S=0.80 E=0.93 VPP=0.94 VPN=0.97
TEPSI Test de desarrollo Psicomotor (Chile) Schapira, 2007 Bedregal P, 2008 Vericat A, 2010	No se encontraron estudios de validación publicados	

BSID: Bayley Scales of Infant Development; SBIS: Stanford Binet Intelligence Scales; MSCA: McCarthy Scales of Children's Abilities; BDI-2: Battelle Development Inventory-2; WPPSI-III: Wechsler Preschool and Primary Scale of Intelligence; KABC: Kaufman Assessment Battery for Children; VABS: Vineland Adaptative Behavior Scale; PLS-IV: Preschool Language Scale-Fourth Edition; WISC: Wechsler Intelligence Scale for Children; EA: Emisiones otoacústicas; PEATC: potenciales evocados auditivos de tallo cerebral; EN: examen neurológico; S: sensibilidad; E: especificidad; VPP: valor predictivo positivo; VPN: valor predictivo negativo.

a cabo y si tiene la capacidad de sustentar su uso. Aunque algunas pruebas tienen un gran número de investigaciones, no significa necesariamente que estas sean óptimas o más rigurosas para detectar retrasos en el desarrollo.

No se encontró evidencia que justifique utilizar un método de aplicación por encima de otro (observación directa, cuestionario para padres o mixto), ya que lo importante en una prueba de tamizaje es su fundamento científico y la

**Cuadro 4.** Ventajas y desventajas de las pruebas de tamizaje según la bibliografía (Continúa en la siguiente página)

<i>Prueba de Tamizaje</i>	<i>Ventajas</i>	<i>Desventajas</i>	<i>Observaciones</i>
Ages & Stages Questionnaires (EUA) Squires & Bricker, 1997 Limbos MM, 2011 Rydz, 2006	Se puede resolver por los padres en casa o en sala de espera (se recomienda sea enviado a casa), su aplicación es de corta duración y de calificación rápida Se puede apoyar el padre con materiales	No se observa de manera directa al niño por personal de salud experimentado	Su sensibilidad y especificidad varía mucho entre un estudio y otro. Va desde una sensibilidad de 0.67 hasta 0.82 y una especificidad tan baja como 0.39
Battelle Developmental Inventory Screening 2nd ed. (EUA) Newborg J, 2005	Requiere poco entrenamiento para su administración Materiales como apoyo Buena sensibilidad y especificidad para el área de comunicación Puede ser modificada para niños con discapacidad	Baja sensibilidad para área motora de 0.64-0.67 y especificidad de 0.74- 0.76	Se utilizaron 3 puntos de corte, con una muestra amplia de niños.
Bayley Scale of Infant and Toddler Development Screening Test (EUA) Aylward GP, 2005	Se apoya con materiales Es de corta duración	Requiere de personal experimentado que pueda evaluar tono muscular y funciones neurológicas básicas	El estudio de validación fue realizado con una muestra significativa de 600 casos, corrigiéndose la edad y con 3 puntos de corte
Brigance Screens-II Glascoe FP, 2005	Corta duración Se puede contestar por padres, por observación directa o ambas formas En la sala de espera o escuelas	No cuenta con materiales de apoyo	En el estudio no explican cuales fueron los puntos de corte
CAT/CLAMS Capute, 1986	Corta duración Buenos materiales Es cuantitativa, establece edad de desarrollo	Considera solo dos dominios del desarrollo (comunicación y resolución de problemas)	
Child Development Inventory (EUA) Doig, et al, 1999 Rydz, et al, 2006	Corto tiempo de aplicación. No cambia la sensibilidad y especificidad dependiendo de nivel socioeconómico de los padres	No tiene materiales de apoyo Tiene en total 300 preguntas y puede durar hasta 30 minutos	Se han realizado estudios de validación para población abierta y de riesgo En un estudio unicamente se evaluaron 36 pacientes comparándolos con estándar de oro (Battelle)
DENVER-II Glascoe, et al, 1992	Es muy fácil de aplicar No requiere entrenamiento intensivo	La sensibilidad y especificidad es baja Tiempo de evaluación de hasta 30 minutos	La muestra sujeta a estudio fue de solo 104 niños de los cuales muy pocos tenían menos de 24 meses A pesar de la reestandarización y utilizando diferentes combinaciones de métodos de calificación no cuenta con una especificidad y sensibilidad adecuadas
Escala de Desarrollo Integral del Niño (Costa Rica) Vericat A, 2010	Propiedades psicométricas adecuadas Observaciones en diferentes niveles socioeconómicos		No se han encontrado datos publicados de validación de esta prueba
EEDP Escala de Evaluación del Desarrollo Psicomotor (Chile) Bedregal P, 2008 Vericat A, 2010	De fácil aplicación Puede ser realizado por personal no especializado	No abarca todos los dominios del desarrollo	No están disponibles en la literatura los estudios de validación

**Cuadro 4.** Ventajas y desventajas de las pruebas de tamizaje según la bibliografía (Continuación)

Prueba de Tamizaje	Ventajas	Desventajas	Observaciones
Neurodesarrollo Pediátrica (Cuba) Guadarrama-Celaya, 2011	Alta sensibilidad y especificidad Puede ser realizada por enfermeras, sin requerir capacitación complicada Corta duración	Es necesario contar con computadora y puede ser necesario capacitación en manejo del software lo que dificulta su aplicación	No se encuentran reportadas las características psicométricas de la prueba
Parents Evaluation of Developmental Status (EUA) Glascoe, et al, 2003 Glascoe, 2001 Limbos MM, Joyce DP; 2011	Se basa en discutir con los padres sus preocupaciones sobre el desarrollo infantil En manos de un evaluador experimentado puede obtenerse información muy valiosa Requiere poco tiempo de administración	Requiere de un evaluador con experiencia para obtener información adecuada	En el estudio de Limbos MM y Joyce DP, se determinó una especificidad muy baja que conlleva a sobre-referencias (1 de cada 3 niños no requerían una valoración)
PRUNAPE (Argentina) Pascucci MC, 2002 y 2004	Diseño semejante al Denver que permite una rápida organización del evaluador sobre la pautas que debe administrar Se apoya con materiales Tiempo de administración breve	La prueba se validó con 78 pautas La pauta "papá-mamá específico" se incorporó luego del proceso de validación	El estudio para su validación con muestra significativa de 839 niños Se utilizaron pruebas diagnósticas para su validación, así como valoraciones por especialistas de audición y visión
Test de desarrollo Psicomotor (Chile) Bedregal P, 2008 Vericat A, 2010	Buen diseño Puede ser utilizado por maestros de preescolar Tiene una versión para ciegos	No evalúa todos los dominios del desarrollo	No tiene datos publicados de validación

metodología del estudio de validación. Por ejemplo, en el estudio de Rydz y colaboradores se encontró que el *Child Development Inventory* presenta una baja sensibilidad con respecto a otros estudios, ya que, únicamente, 5 de 31 niños fallaron en *Battelle Developmental Inventory* (comparador), por lo que sería necesario analizar una muestra más grande.

Con respecto a la puntuación de las escalas existen sistemas de calificación cuantitativos (con un sistema de puntuación por ítem) o cualitativos (con una clasificación categórica, como falla/aprueba, presente/ausente, sí/no, entre otros). No se encontró una tendencia de mayor confiabilidad con respecto a algún método de calificación en las distintas pruebas de tamizaje incluidas en este estudio, ya que, por ejemplo, la PRUNAPE y la Battelle tienen sistemas de calificación diferentes. Sin embargo, fueron las dos pruebas con mejor sensibilidad y especificidad encontradas.

Con respecto a los estudios incluidos en esta revisión se encontró que muy pocos cumplen las condiciones óptimas para apoyar su confiabilidad, validez y utilidad, ya que algunos presentan limitaciones en cuanto a su diseño. Pocos

estudios emplean un diseño cuasiexperimental o de asignación aleatoria, y esto resulta uno de los principales problemas para su confiabilidad. Además, la exactitud en la administración de la prueba es crucial, aunque existen limitantes tanto en el conocimiento de la prueba y experiencia de quien lo aplica, como el tiempo y lugar en que se administra. Esto condiciona una variabilidad en los resultados tan importante como tener una sensibilidad de 0.67 a 0.90 y especificidad de 0.39 a 0.95, como en el caso de los estudios de *Ages & Stages Questionnaires*, donde en algunos casos el cuestionario se aplicó a los padres en la sala de espera, con un tiempo de 15 minutos, y en otros se les envió a casa, dando tiempo a una mayor fidelidad en los resultados.

El tamaño de la muestra en algunos estudios no es suficiente. Resulta especialmente problemático cuando se calcula la especificidad y sensibilidad de la prueba, debido a que la cantidad de pacientes con una afección de desarrollo puede ser limitada. En otros casos, el problema radica en que no se incluye un número similar de pacientes por grupo de edad, como en el caso de Denver, donde además de tener una muestra pequeña (104) muy pocos eran menores de 24 meses.

Es de suma importancia determinar puntos de corte adecuados para calibrar la herramienta, de manera que los resultados sean precisos. Aquellas pruebas que tienen baja especificidad producen un gran número de falsos positivos que puede resultar en una sobrerreferencia diagnóstica. El problema opuesto conlleva a no referir y no detectar a aquellos niños con un problema en el desarrollo y, por consiguiente, no intervenir tempranamente.

Con base en los resultados que se observaron, se concluye que entre el período de 1980-2012, que comprende la revisión sistemática, se encontraron 13 pruebas de tamizaje del neurodesarrollo en Estados Unidos y Latinoamérica para su aplicación en niños menores de 5 años de edad.

*Battelle Development Inventory Screening* y PRUNAPE fueron las dos pruebas de tamizaje que se encontraron con mayor sensibilidad y especificidad en la validez de criterio. Cabe destacar que la evidencia de su validación ofrecida por las publicaciones es de alta calidad metodológica, lo que las confirma como herramientas confiables para la detección de alteraciones en el neurodesarrollo.

Si bien, existen diferentes métodos de administración, calificación y criterios de normalidad y anormalidad, no se encontró evidencia científica que justifique uno u otro sistema de la prueba, ya que lo importante es el fundamento científico bajo el cual se realiza la validación de la prueba.

De las pruebas de tamizaje incluidas, no se encontró ningún estudio de validación en México, por lo que consideramos de suma importancia contar con una herramienta validada, preferentemente propia, para ser aplicada en nuestra población, y con esto implementar intervenciones tempranas de manera sistemática, dirigida y con fundamento científico. Lo anterior para lograr el potencial máximo de desarrollo y evitar los factores causantes de pobreza intergeneracional.

*Autor de correspondencia:* Dra. Beatriz Romo Pardo  
Correo electrónico: beatrizromo\_doc@hotmail.com

## REFERENCIAS

1. Council on Children With Disabilities; Section on Developmental Behavioral Pediatrics; Bright Futures Steering Committee; Medical Home Initiatives for Children With Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. *Pediatrics* 2006;118:405-420.
2. Grantham-McGregor S, Cheung YB, Cueto S, Glewwe P, Richter L, Strupp B, et al. Developmental potential in the first 5 years for children in developing countries. *Lancet* 2007;369:60-70.
3. Hamilton S. Screening for developmental delay: reliable, easy-to-use tools. *J Fam Pract* 2006;55:415-422.
4. Rydz D, Srour M, Oskoui M, Marget N, Shiller M, Birnbaum R, et al. Screening for developmental delay in the setting of a community pediatric clinic: a prospective assessment of parent-report questionnaires. *Pediatrics* 2006;118:e1178-e1186.
5. Mackrides PS, Ryherd SJ. Screening for developmental delay. *Am Fam Physician* 2011;84:544-549.
6. Rydz D, Shevell MI, Majnemer A, Oskoui M. Developmental screening. *J Child Neurol* 2005;20:4-21.
7. Glen P, Aylward T. S. Measurement and psychometric considerations. En: Wolraich ML, Drotar DD, Dworkin PH, Perrin EC, eds. *Developmental-Behavioral Pediatrics: Evidence and Practice*. Philadelphia: Mosby Elsevier; 2008. pp. 123-201.
8. Squires J, Bricker D, Potter L. Revision of a parent-completed developmental screening tool: Ages and Stages Questionnaires. *J Pediatr Psychol* 1997;22:313-328.
9. Squires J, Twombly E, Bricker D, Potter L. *ASQ-3™ User's Guide*. Baltimore, MD: Paul H. Brookes Publishing Co., Inc; 2009.
10. Schonhaut BL, Salinas AP, Armijo RI, Schönstedt GM, Álvarez LJ, Manríquez OM. Validación de un cuestionario autoadministrado para la evaluación del desarrollo psicomotor. *Rev Chil Pediatr* 2009;80:513-519.
11. Limbos MM, Joyce DP. Comparison of the ASQ and PEDS in screening for developmental delay in children presenting for primary care. *J Dev Behav Pediatr* 2011;32:499-511.
12. Newborg J. Development and standardization. En: Newborg J, ed. *Battelle Developmental Inventory*. Itaska, IL: Riverside Publishing; 2004. pp. 95-148.
13. Aylward GP. The Bayley Infant Neurodevelopmental Screener (BINS): different test and different purpose. En: Weiss LG, Oakland T, Aylward G, eds. *Bayley III Clinical Use and Interpretation*. USA: Academic Press; 2010. pp. 201-233.
14. Glascoe FP. The Brigance Infant and Toddler Screen: standardization and validation. *J Dev Behav Pediatr* 2002;23:145-150.
15. Capute AJ, Shapiro BK, Watchel RC, Gunther VA, Palmer FB. The Clinical Linguistic and Auditory Milestone Scale (CLAMS). Identification of cognitive defects in motor-delayed children. *Am J Dis Child* 1986;140:694-698.
16. Doig KB, Macias MM, Saylor CF, Craver JR, Ingram PE. The Child Development Inventory: a developmental outcome measure for follow-up of the high-risk infant. *J Pediatr* 1999;135:358-362.
17. Glascoe FP, Byrne KE, Ashford LG, Johnson KL, Chang B, Strickland B. Accuracy of the Denver-II in developmental screening. *Pediatrics* 1992;89:1221-1225.
18. Schapira IT. Comentarios y aportes sobre desarrollo e inteligencia sensorio-motriz en lactantes. Análisis de herramientas de evaluación de uso frecuente. Actualización bibliográfica. *Rev Hosp Mat Infanz Ramón Sardá* 2007;26:21-27.
19. Bedregal P. Instrumentos de medición del desarrollo en Chile. *Rev Chil Pediatr* 2008;79(suppl 1):32-36.
20. Vericat A, Orden AB. Herramientas de screening del desarrollo psicomotor en América. *Rev Chil Pediatr* 2010;81:391-401.
21. Guadarrama-Celaya F, Otero-Ojeda GA, Pliego-Rivero B, Porcayo-Mercado MR, Garcell JR, Pérez-Ábalo MC. Screening

- of neurodevelopmental delays in four communities of Mexico and Cuba. *Public Health Nurs* 2012;29:105-115.
22. Pascucci MC, Lejarraga H, Kelmansky D, Álvarez M, Boullón M, Breiter P, et al. Validación de la prueba nacional de pesquisa de trastornos de desarrollo psicomotor en niños menores de 6 años. *Arch Argen Pediatr* 2002;100:374-384.
23. Pascucci MC, Lejarraga H, Kelmansky D, Álvarez M, Boullón M, Breiter P, et al. Validación de la prueba nacional de pesquisa de trastornos de desarrollo psicomotor en niños menores de 6 años. *Arch Pediatr Urug* 2004;75:79-90.