



Jueves 19 de octubre de 2006

Mesa redonda:

“Revisión de los artículos pediátricos de mayor impacto y de conocimiento imprescindible para el pediatra de Atención Primaria”

Moderador:

Juan Ruiz-Canela Cáceres

Pediatra, CS de Torreblanca-Distrito, Sevilla.

■ **Artículos sobre intervenciones terapéuticas**

Paz González Rodríguez

Pediatra, CS Manzanares el Real. Área 5. IMSALUD. Madrid.

Javier González de Dios

Pediatra, Departamento de Pediatría. Hospital Universitario San Juan. Universidad Miguel Hernández. Alicante.

■ **Estudios sobre prevención. Vacunación frente a rotavirus**

Pedro Martín Muñoz

Pediatra, CS La Plata, Distrito AP. Sevilla.

■ **Estudios sobre pruebas diagnósticas**

Carlos Ochoa Sangrador

Servicio de Pediatría y Unidad de Apoyo a la Investigación. Hospital Virgen de la Concha. Zamora.

Textos disponibles en

www.aepap.org

¿Cómo citar este artículo?

Ochoa Sangrador C. Estudios sobre pruebas diagnósticas. En: AEPap ed. Curso de Actualización: Pediatría 2006. Madrid: Exlibris Ediciones; 2006. p. 43-55.



Estudios sobre pruebas diagnósticas

Carlos Ochoa Sangrador

Servicio de Pediatría y Unidad de Apoyo a la Investigación.

Hospital Virgen de la Concha. Zamora.

cochoas@meditex.es

RESUMEN

Consideramos prueba diagnóstica a cualquier procedimiento realizado para confirmar o descartar un diagnóstico o para incrementar o disminuir su verosimilitud. Si queremos tomar las mejores decisiones en la elección y aplicación de pruebas diagnósticas para nuestros pacientes, tenemos que estar preparados para integrar nuestros conocimientos y experiencia previos con la información científica de los trabajos que continuamente vienen incorporándose a la literatura médica. Repasamos, en primer lugar, los parámetros que describen la validez y la fiabilidad de las pruebas diagnósticas y la metodología de evaluación de la validez, la relevancia y la aplicabilidad de los estudios donde son estimados dichos parámetros. Analizamos a continuación los trabajos publicados recientemente sobre pruebas diagnósticas en la infancia. De todos los temas estudiados destacan, por su relevancia clínica, el cuestionado papel del reflujo vesicoureteral como factor de riesgo de daño renal crónico (por el que se cuestiona la realización de la cistografía), la utilidad de las pruebas de imagen y de los modelos predictivos clínicos en el diagnóstico de apendicitis aguda, la efectividad del cribado universal de hipoacusia neonatal y de displasia de desarrollo de caderas, la validez de las pruebas diagnósticas no invasivas de gastritis ulcus por *Helicobacter pylori* y la influencia de la técnica de recogida de orina en el estudio de infección urinaria. Asimismo, presentamos de forma resumida los indicadores de validez, cocientes de probabilidades y probabilidades posprueba de otras pruebas diagnósticas. Planteamos finalmente la importancia de los cálculos ajustados a cada escenario clínico y la consideración del impacto sobre el comportamiento diagnóstico o terapéutico en la valoración de la verdadera utilidad clínica de cada prueba diagnóstica.

INTRODUCCIÓN

El diagnóstico médico es un proceso dinámico en el que se intentan tomar decisiones idóneas en presencia de incertidumbre. Desde un punto de vista funcional, consideramos prueba diagnóstica a cualquier procedimiento realizado para confirmar o descartar un diagnóstico o incrementar o disminuir su verosimilitud. La utilidad de una prueba diagnóstica depende fundamentalmente de su validez y de su fiabilidad, pero también de su rendimiento clínico y de su coste. Si evaluamos nuestra práctica clínica, observaremos que a menudo empleamos procedimientos diagnósticos de los que desconocemos estos parámetros. Para conocer esta información debemos recurrir a los estudios de evaluación de pruebas diagnósticas publicados. Sin embargo, en muchos de estos estudios no siempre encontramos información válida, relevante o aplicable a nuestro entorno clínico¹⁻⁴.

En la literatura científica aparecen continuamente estudios de evaluación de pruebas diagnósticas. Estos estudios pueden corresponder a nuevas herramientas diagnósticas, desarrolladas como resultado de avances tecnológicos, que tratan de mejorar las características operativas de las pruebas ya disponibles, o bien mejorar su laboriosidad, rapidez, accesibilidad, efectos adversos o coste. En otras ocasiones, estos estudios actualizan la información de pruebas ya existentes, comparando diversas opciones o explorando su rendimiento en distintos escenarios clínicos o con patrones de referencia alternativos. Otros trabajos revisan la utilidad clínica de signos o síntomas clínicos ya empleados en la rutina diaria que, analizados de forma sistemática, como modelos predictivos, pueden ser aplicables en el cálculo de probabilidades del proceso diagnóstico.

Si queremos tomar las mejores decisiones en la elección y aplicación de pruebas diagnósticas para nuestros pacientes, tenemos que estar preparados para integrar nuestros conocimientos y experiencia previos con la información científica de los trabajos que continuamente vienen incorporándose a la literatura médica. Para ello, necesitamos estar familiarizados con los parámetros

que describen la validez y la fiabilidad de las pruebas diagnósticas y aprender a evaluar la validez, la relevancia y la aplicabilidad de los estudios donde son estimados^{5,6}. Antes de revisar los trabajos publicados recientemente sobre pruebas diagnósticas en la infancia, repasaremos brevemente estos aspectos.

VALIDEZ Y FIABILIDAD DE LAS PRUEBAS DIAGNÓSTICAS

El concepto de validez se refiere a la capacidad de una prueba para medir lo que realmente queremos medir. La validez se evalúa comparando los resultados de la prueba con los de un patrón de referencia (*gold-standard*), que identifica el diagnóstico verdadero. Para pruebas con resultados dicotómicos (ej.: presencia-ausencia de enfermedad), la evaluación se concreta en distintos indicadores de validez: sensibilidad (probabilidad de que la prueba dé positiva si la enfermedad o condición de estudio está presente), especificidad (probabilidad de que la prueba dé negativa si la enfermedad está ausente) y valores predictivos positivo (probabilidad de tener la enfermedad si la prueba ha sido positiva) y negativo (probabilidad de no tener la enfermedad si la prueba ha sido negativa) (tabla I).

La sensibilidad y la especificidad son características intrínsecas de la prueba diagnóstica que informan de su validez antes de conocer el resultado. Una vez conocido el resultado de la prueba, son los valores predictivos los que informan sobre la probabilidad de la enfermedad. Sin embargo, los valores predictivos dependen de la prevalencia de enfermedad en el estudio original y por ello no son directamente aplicables a escenarios con probabilidades preprueba distintas. Para conocer la probabilidad de enfermedad (posprueba) en escenarios con probabilidades preprueba distintas debemos usar los cocientes de probabilidades (probabilidad de un determinado resultado de la prueba en presencia de enfermedad dividida por la probabilidad de ese mismo resultado en ausencia de enfermedad) que son índices resumen de la sensibilidad y la especificidad, independientes de la probabilidad preprueba en el estudio original.

Tabla I. Tabla de contingencia de la evaluación de una prueba diagnóstica. Características operativas de la prueba

		Patrón de referencia		
		+	-	
Prueba diagnóstica	+	Verdaderos positivos (a)	Falsos positivos (b)	a + b
	-	Falsos negativos (c)	Verdaderos negativos (d)	c + d
		a + c	b + d	Total = a + b + c + d

Claves:

a: Verdaderos positivos (VP): enfermos con la prueba positiva
 b: Falsos positivos (FP): no enfermos con la prueba positiva
 c: Falsos negativos (FN): enfermos con la prueba negativa
 d: Verdaderos negativos (VN): no enfermos con la prueba negativa
 a + c: Casos con patrón de referencia positivo (enfermos)
 b + d: Casos con patrón de referencia negativo (no enfermos)
 a + b: Casos con la prueba diagnóstica positiva
 c + d: Casos con la prueba diagnóstica negativa
 Sensibilidad (Se) = $a / (a + c)$
 Especificidad (Es) = $d / (b + d)$
 Valor predictivo positivo (VPP) = $a / (a + b)$
 Valor predictivo negativo (VPN) = $d / (c + d)$
 Cociente de probabilidad (CP) positivo = $Se / (1 - Es)$
 Cociente de probabilidad negativo = $(1 - Se) / Es$
 Probabilidad preprueba (Ppre) = $(a + c) / (a + b + c + d)$
 Odds preprueba (odds pre) = $Ppre / (1 - Ppre)$
 Odds posprueba (odds post) = $CP \text{ positivo} \times \text{odds pre}$
 Probabilidad postprueba (Ppost) = $\text{odds post} / (1 + \text{odds post})$

Para poder operar con los CP en el cálculo de probabilidades, éstas deben transformarse en ventajas (*odds*). Las ventajas u odds se calculan dividiendo las probabilidades por sus complementarios ($P / [1 - P]$). Los pasos que hay que seguir en el cálculo de la probabilidad posprueba son: 1) transformar la probabilidad preprueba en odds preprueba, 2) multiplicar la odds preprueba por el CP del resultado encontrado (positivo o negativo habitualmente) para obtener la odds posprueba, 3) transformar la odds postprueba en probabilidad (probabilidad = $\text{odds} / [1 + \text{odds}]$). Estos cálculos se simplifican recurriendo a nomogramas (de Fagan) o a aplicaciones informáticas. Es importante recordar que la utilización de una prueba diagnóstica no permite afirmar o rechazar absolutamente un diagnóstico por muy buenas que sean sus capacidades operativas, sólo obtendremos un incremen-

to o un descenso de su probabilidad (de preprueba a posprueba).

Todos estos estimadores de validez pueden aplicarse a pruebas con resultados discretos con más de dos categorías e incluso a resultados expresados en variables continuas. En este caso podemos establecer puntos de corte para convertir el resultado en una variable discreta con dos o más categorías. Otra alternativa que permite explorar la capacidad diagnóstica de una prueba en sus distintos valores son las curvas ROC (iniciales del término inglés original *Receiver Operating Characteristics*), con las que podemos conocer su validez global y seleccionar el punto o puntos de corte más adecuados. Es frecuente encontrar trabajos en los que se selecciona un punto de corte desplazado de su

punto medio, con la intención de maximizar sensibilidad a expensas de reducir especificidad, o viceversa; en estos casos, podemos estar limitando la aplicabilidad de la prueba ya que dichos valores sólo le serán útiles a una parte de los pacientes.

Es preciso tener en cuenta que la información de que disponemos sobre la validez de las pruebas diagnósticas procede de estudios realizados en muestras de población. Por tanto, las estimaciones obtenidas en dichos estudios están sujetas a variabilidad aleatoria (por lo que les son aplicables intervalos de confianza) y si los estudios han sido diseñados incorrectamente, a sesgos.

La fiabilidad de una prueba viene determinada por la estabilidad de sus mediciones cuando se repite en condiciones similares. La variabilidad de las mediciones va a estar influida por múltiples factores que interesa conocer y controlar. Entre ellos, tiene especial importancia distinguir las variaciones de interpretación intraobservador e interobservador. La fiabilidad puede ser evaluada para resultados discretos nominales mediante el índice kappa, para resultados discretos ordinales mediante el índice kappa ponderado y para resultados continuos mediante el coeficiente de correlación intraclase y el método de Bland-Altman.

EVALUACIÓN DE UNA PRUEBA DIAGNÓSTICA

A la hora de analizar la evidencia que un determinado estudio aporta sobre la validez y la aplicabilidad de una prueba diagnóstica, debemos plantearnos diversas cuestiones, estructuradas en tres pasos, según el esquema propuesto por el *Evidence Based Medicine Working Group*^{7,8}:

- ¿Son válidos sus resultados?
- ¿Cuáles son los resultados?
- ¿Son aplicables en tu medio?

¿Son válidos sus resultados?

Para poder valorar la validez de una prueba diagnóstica es preciso comparar sus resultados con los de un patrón de referencia en una serie de pacientes. El patrón de referencia empleado tiene que contar con una validez contrastada o, al menos, aceptada por consenso. La utilización de un patrón de referencia defectuoso puede introducir sesgos en las estimaciones de validez de la prueba diagnóstica.

En relación con el patrón de referencia, resulta también importante considerar si es capaz de clasificar el estado de enfermedad en todas las observaciones. En el caso de que existan observaciones con un diagnóstico indeterminado, si éstas son excluidas del análisis, se producirán estimaciones sesgadas de las características operativas de la prueba diagnóstica. Este sesgo, conocido como sesgo por exclusión de indeterminados, ocasiona habitualmente sobrestimaciones de la sensibilidad y de la especificidad.

Si aceptamos la validez del patrón de referencia, la siguiente cuestión que hay que tener en cuenta es si tanto la prueba diagnóstica como el patrón de referencia han sido realizados de forma independiente. Cuando no se realizan de forma independiente, puede existir un sesgo de revisión si el resultado de una prueba es susceptible de interpretación subjetiva y se ve influida por el conocimiento del diagnóstico o de las características clínicas del paciente. Para poder garantizar la validez de las estimaciones, deberían realizarse de forma ciega la prueba diagnóstica y el patrón de referencia.

Otro aspecto importante que condiciona la validez del estudio es la inclusión en éste de un adecuado espectro de pacientes, similar al que nos encontramos en nuestra práctica clínica. Para valorar esta cuestión es preciso que los criterios de selección y las características clínicas y epidemiológicas de la muestra analizada estén claramente presentados.

El diseño del estudio debe tratar de garantizar que en la muestra no se haya excluido a pacientes en función

del resultado de la prueba o de la existencia de mayor o menor riesgo de enfermedad. Incurriremos en un sesgo de verificación diagnóstica cuando la probabilidad de que se les realice el patrón de referencia sea menor entre los sujetos con la prueba diagnóstica negativa y, por tanto, sea menos probable que éstos entren en el estudio.

El último criterio que hay que valorar para juzgar la validez del estudio es si se describen los métodos con suficiente detalle como para permitir su reproducción. Esta descripción debe incluir todos los aspectos de la preparación de pacientes, la realización de la prueba y su interpretación.

Si después de considerar todos estos aspectos hemos decidido que el estudio es suficientemente válido, procederemos a examinar las propiedades de la prueba diagnóstica.

¿Cuáles son los resultados?

El punto de partida del proceso diagnóstico es habitualmente un paciente, con unas características de gravedad y comorbilidad concretas, que le confieren una probabilidad determinada de tener la entidad que hay que diagnosticar (probabilidad preprueba). El objetivo de la realización de la prueba diagnóstica es, una vez conocido el resultado, modificar esa probabilidad hasta obtener una probabilidad posprueba. La magnitud y la dirección de ese cambio van a depender de las características operativas de la prueba diagnóstica, pero en todo caso debemos tener en cuenta que el punto de partida, la probabilidad preprueba, va a resultar muy importante en ese proceso.

Consideremos el escenario diagnóstico más simple, en el que tanto el patrón de referencia como la prueba diagnóstica clasifican a los pacientes en dos grupos en función de la presencia o ausencia de un síntoma, signo o enfermedad. Utilizando los CP se pueden calcular las probabilidades posprueba (valores predictivos) a partir de la probabilidad preprueba de cada paciente individual (tabla I), que habitualmente no es la misma

que la existente en los estudios publicados. La relevancia de la prueba va a depender de cuánto sea capaz de incrementar o disminuir esa probabilidad una vez aplicados los CP positivo o negativo. Generalmente, los CP positivos cercanos a 10 y negativos cercanos a 0,1 resultan clínicamente relevantes, aunque finalmente serán las probabilidades posprueba ajustadas a mi paciente las que indicarán el grado de relevancia diagnóstica.

Cuando la prueba que se evalúa se mide en una escala cuantitativa, la sensibilidad y la especificidad varían en función del punto de corte que se utilice para separar los valores normales de los anormales. En estos casos, los resultados pueden representarse gráficamente como una curva ROC, que permite conocer las características de la prueba según diferentes puntos de corte y que puede utilizarse para elegir el más adecuado. La relevancia global de la prueba va a depender del tamaño del área bajo la curva; comparando las áreas de distintas pruebas podremos conocer la que mayor capacidad diagnóstica tiene.

Al igual que en otros tipos de estudios, la valoración de la validez de las pruebas diagnósticas se hace sobre muestras, por lo que los resultados obtenidos son sólo estimaciones puntuales, sujetas a variabilidad aleatoria y, por tanto, deben proporcionarse con sus intervalos de confianza. Estos intervalos de confianza tendrán que aplicarse en el cálculo de la probabilidad posprueba para poder juzgar la utilidad de la prueba diagnóstica.

¿Son aplicables en tu medio?

A la hora de valorar la aplicabilidad de la prueba diagnóstica en nuestros pacientes tenemos que considerar si la prueba es suficientemente reproducible, independientemente del ámbito y las condiciones de aplicación y de la persona que la interprete. Por ello, resulta importante que los trabajos publicados incluyan información expresa sobre la reproducibilidad de la prueba y los criterios de interpretación de ésta.

Otro aspecto que hay que tener en cuenta es si resulta razonable asumir que las características operativas

de la prueba diagnóstica, estimadas en el estudio, van a ser similares en nuestros pacientes. Si el ámbito en el que se ha evaluado la prueba es similar al nuestro, nuestros pacientes cumplen los criterios de inclusión del estudio y no violan los de exclusión, parece sensato aceptar la aplicabilidad de los resultados. Si el espectro de pacientes incluido en el estudio es diferente al nuestro, la decisión deberá tomarse con cautela.

La utilidad clínica de la prueba depende también de la repercusión que tengan sobre nuestra actitud diagnóstica y terapéutica. En ocasiones la probabilidad preprueba de nuestro paciente será tan baja que, independientemente de cuál sea el resultado de la prueba, la probabilidad posprueba será igualmente baja y, por tanto, no merecerá la pena llevarla a cabo (umbral diagnóstico). En el otro extremo, si la probabilidad preprueba es muy elevada, su resultado no va a modificar nuestra decisión de tratar (umbral terapéutico), por lo que a veces podremos obviarla. En la zona intermedia, donde más interés debería tener la prueba, el grado de información que nos aporta dependerá de la magnitud de los cocientes de probabilidades; valores cercanos a 1 resultarán poco útiles, mientras que valores lejanos modificarán de forma importante las probabilidades posprueba y su rendimiento diagnóstico.

Pero el criterio último de la utilidad de una prueba, al margen de que ésta ofrezca información diagnóstica no disponible previamente o de que modifique nuestro comportamiento clínico, es si el paciente obtiene algún beneficio. Existen escenarios en los que la prueba diagnóstica no resulta coste-efectiva, conlleva riesgos o conduce a decisiones terapéuticas sin repercusión sobre el paciente.

ESTUDIOS DE EVALUACIÓN DE PRUEBAS DIAGNÓSTICAS RECIENTEMENTE PUBLICADOS

Para seleccionar los principales estudios publicados en la literatura científica en el último año sobre evaluación de pruebas diagnósticas de interés clínico en pediatría, hemos revisado la base de datos de Medline utilizando

un filtro sensible para estudios de pruebas diagnósticas junto con descriptores relacionados con la infancia y pediatría. Asimismo, hemos revisado las últimas actualizaciones incluidas en varias publicaciones electrónicas (Up To-Date, Nelson Online) y los números publicados hasta el momento en *Evidencias en Pediatría*⁹⁻¹².

Podemos agrupar los trabajos publicados por temas de interés, destacando:

- Apendicitis aguda (utilidad de la tomografía computarizada^{13,14}, validez de modelos predictivos¹⁵, impacto de las pruebas de imagen^{14,16}).
- Reflujo vesicoureteral (significación clínica de su diagnóstico^{17,22}, utilidad de la cistografía PIC²³, validez de la procalcitonina²⁴ e interleucina 8²⁵, validez de modelos predictivos clínicos²⁶).
- Hipoacusia neonatal (utilidad del cribado universal²⁷⁻³³, comparación de estrategias con otonemias evocadas y potenciales auditivos automáticos³⁴).
- Gastritis ulcus por *Helicobacter pylori* (validez de las pruebas no invasivas³⁵⁻³⁷).
- Infección neonatal (validez diagnóstica de procalcitonina y proteína C reactiva³⁸, validez de los parámetros de líquido cefalorraquídeo en la meningitis neonatal³⁹).
- Displasia de desarrollo de cadera (utilidad del cribado ecográfico⁴⁰⁻⁴²).
- Alergia alimenticia (nuevas pruebas diagnósticas de sensibilización atópica⁴³).
- Enfermedad celíaca (validez de la determinación de anticuerpos anti gliadina y antitransglutaminasa en heces⁴⁴).
- Infección urinaria (comparación de técnicas de recogida de orina^{45,46}).

- Estrabismo (validez de diversos instrumentos de cribado^{47,48}).
- Traumatismo craneal (utilidad de la radiografía de cráneo⁴⁹).
- Faringitis estreptocócica (validez de modelos clínicos predictivos⁵⁰).
- Fibrosis quística de páncreas (comparación de pruebas de cribado neonatal⁵¹).
- Tuberculosis (nuevos métodos de diagnóstico serológico de infección latente^{52,53}, utilidad de la radiografía de tórax para el diagnóstico de adenopatías mediastínicas⁵⁴).
- Ferropenia (validez del contenido de hemoglobina reticulocitaria⁵⁵).
- Síndrome de QT largo congénito (validez de la prueba de estrés con epinefrina⁵⁶).
- Encefalopatía aguda con fallo multiorgánico (valor predictivo del citocromo c sérico y las citoquinas⁵⁷).
- Infección bacteriana grave en el lactante febril (valor predictivo de la leucocitosis extrema⁵⁸).
- Crisis epilépticas (utilidad diagnóstica de la apertura de ojos durante la crisis⁵⁹).
- Deficiencia de hormona de crecimiento (validez de la determinación de IGF e IGFBP3⁶⁰).

Ante la imposibilidad de tratar todos estos temas nos centramos en aquéllos de especial interés por su repercusión clínica. Como complemento, en la tabla II presentamos una relación de los principales resultados de algunos trabajos, en los que hemos podido extraer los indicadores de validez, CP y probabilidades preprueba y posprueba de varias pruebas diagnósticas.

El tema que mayor debate ha suscitado en el último año es el controvertido papel del reflujo vesicoureteral como factor de riesgo de daño renal crónico⁶¹, por el que se cuestiona la realización de pruebas diagnósticas y tratamientos que usamos rutinariamente (cistografía y profilaxis antibiótica). En las últimas décadas el manejo de la infección urinaria en la infancia ha girado en torno al diagnóstico de reflujo vesicoureteral; su diagnóstico implicaba la asunción de un riesgo de desarrollo de daño renal (nefropatía por reflujo) y por consiguiente se consideraba necesaria la realización y la repetición de cistografías, así como la instauración de profilaxis antibiótica. En el momento actual existe información suficiente que sugiere que el reflujo en la primera infancia es más frecuente de lo que se creía⁶², que el diagnóstico de reflujo de bajo grado y su persistencia tiene escasa repercusión sobre el desarrollo de cicatrices renales²⁰, que la profilaxis antibiótica tiene escaso impacto¹⁷ y que el daño renal preexistente en el momento del diagnóstico es el factor que más condiciona el pronóstico⁶¹. Como consecuencia, el empleo rutinario de la cistografía y de la profilaxis antibiótica no parece justificado. Asimismo, considerando la escasa relevancia clínica del reflujo vesicoureteral, se entiende que diversos estudios hayan observado que tanto los modelos predictivos clínicos²⁶, como la procalcitonina²⁴ o la interleucina 8 urinaria²⁵, ofrecen un bajo rendimiento en su diagnóstico.

Otro tema analizado en diversos trabajos es el diagnóstico de la apendicitis aguda. En los últimos años se ha generalizado el uso de técnicas de imagen. Así, es cada vez más habitual la realización de ecografías, que en ocasiones nos ofrecen imágenes de dudosa interpretación o que no concuerdan con la clínica, por lo que se llega a recurrir a la realización de tomografías computerizadas (TC), especialmente en algunos países. Dos trabajos recientes analizan el impacto del empleo de la TC^{14,16}, aunque con algunas limitaciones que dificultan la interpretación de sus resultados. Desde la introducción de la TC, parece existir una tendencia en el tiempo a la disminución de las apendicectomías "blancas"¹⁴, aunque su uso implica un mayor coste económico y un cierto retraso diagnóstico y quirúrgico¹⁶. La

Tabla II. Indicadores de validez (sensibilidad y especificidad), cocientes de probabilidades (CP) y probabilidades posprueba de varias pruebas diagnósticas

Diagnóstico Prueba ^{cita}	Ppre	Se	Es	CP+	CP-	Ppost+	Ppost-	Comentario
Apendicitis aguda								
• Modelo clínico; Escala ¹⁵	31%	96%	36%	1,44	0,11	40%	4%	Sólo útil en el 25% de casos con bajo riesgo
• Modelo clínico; RDC ¹⁵	31%	98%	32%	1,44	0,06	39%	2%	
Deficiencia de GH								
• IGF-I ⁶⁰	55%	69%	81%	3,63	0,38	82%	32%	Muestra de casos y controles Ppre no aplicable PC mediana grupos de edad
• IGFBP-3 ⁶⁰	55%	27%	100%	-	0,73	100%	47%	
Reflujo vesicoureteral								
• Interleucina 8 / creatinina orina > 5 pg / μmol ²⁵	41%	88%	69%	2,83	0,17	66%	11%	Ppre no aplicable
• Procalcitonina > 0,5ng / ml ²⁴	25%	85%	44%	1,51	0,34	34%	10%	
• Modelo clínico ²⁶	25%	100%	3%	1,03	0,00	26%	0%	
Meningitis neonatal								
• Leucocitos LCR > 0 / mmc ³⁹	1%	97%	11%	1,08	0,27	1,1%	0,3%	PR cultivo LCR
• Leucocitos LCR >20 / mmc ³⁹	1%	79%	81%	4,16	0,24	4,0%	0,3%	
• Glucosa < 20 mg / dl ³⁹	1%	44%	98%	22,00	0,05	18,2%	0,6%	
• Proteínas > 90 mg / dl ³⁹	1%	76%	63%	2,05	0,49	2,0%	0,4%	
Alergia a proteínas vacunas								
• Test cutáneo Diallerstest ¹³	60%	76%	94%	12,66	0,25	95%	28%	PR prueba provocación oral
Gastritis-úlcus por <i>H.pylori</i>								
• IgG sérica anti- <i>H.pylori</i> -1 ³⁵	56%	82%	70%	2,73	0,25	78%	25%	PR histología y test ureasa en biopsia endoscópica 1 = Bioelisa Helicobacter 2 = Test no comercializado PC ajustados a la población
• IgA sérica anti- <i>H.pylori</i> -1 ³⁵	56%	85%	53%	1,80	0,28	70%	26%	
• IgG sérica anti- <i>H.pylori</i> -2 ³⁵	56%	91%	64%	2,52	0,14	76%	15%	
• IgA sérica anti- <i>H.pylori</i> -2 ³⁵	56%	85%	55%	1,88	0,27	71%	26%	
• Ag <i>H.pylori</i> en heces ³⁷	37%	97%	98%	48,50	0,03	97%	2%	
• Ag <i>H.pylori</i> en heces ³⁶	40%	80%	93%	11,42	0,21	88%	13%	
• Urea-C ¹³ en aire espirado ³⁶	40%	96%	97%	32,00	0,04	96%	3%	
• Ag <i>H.pylori</i> en heces ³⁷	37%	97%	98%	48,50	0,03	97%	2%	
Daño renal tras primera infección urinaria sin hallazgos ecográficos								* PR DMSA a los 3-6 meses Cistografía convencional en niños, isotópica en niñas
• Reflujo vesicoureteral ²⁰	3,7%	50%	90%	4,95	0,56	16%	2%	
Infeción urinaria								
• Tira reactiva orina bolsa ⁴⁵	27%	85%	62%	2,23	0,24	52%	7%	PR cultivo orina catéter Lectura automática de la tira (leucocituria y/o nitritos)
• Tira reactiva orina catéter ⁴⁵	27%	71%	97%	23,66	0,29	90%	10%	
Faringitis estreptocócica								
• RDC OMS (Brasil) ⁵⁰	25%	9%	96%	2,02	0,95	43%	24%	PR cultivo frotis faríngeo Modelos predictivos clínicos
• RDC OMS (Croacia) ⁵⁰	42%	6%	94%	1	1	42%	42%	
• RDC OMS (Egipto) ⁵⁰	28%	4%	97%	1,36	0,98	34%	28%	

Ppre: probabilidad preprueba en la muestra del estudio; Se: sensibilidad; Es: especificidad; CP+: cociente de probabilidades positivo; CP-: cociente de probabilidades negativo; Ppost+: probabilidad posprueba positiva; Post-: probabilidad posprueba negativa; PR: patrón de referencia; PC: punto de corte; RDC: regla de decisión clínica; OMS: Organización Mundial de la Salud.

inclusión en estos estudios de pacientes con clínica muy clara de apendicitis, que son intervenidos sin recurrir a pruebas de imagen, origina que, al comparar el porcentaje apendicectomías “blancas” de los pacientes con y sin técnicas de imagen, no se encuentren diferencias significativas¹⁴, o que incluso ese porcentaje pueda ser mayor entre los que se emplean pruebas de imagen¹⁶. Por el mismo motivo, resulta difícil valorar el impacto de la TC sobre otros parámetros, como el riesgo de apendicitis complicadas. Considerando la información ofrecida por estos estudios, no parece justificado el uso rutinario de la TC en el diagnóstico de apendicitis, aunque podamos recurrir a esta técnica en casos seleccionados.

Además del papel de las pruebas de imagen en el diagnóstico de apendicitis aguda, también sigue despertando interés el estudio de otros recursos diagnósticos clásicos: síntomas clínicos, signos exploratorios y fórmula leucocitaria en sangre. Un trabajo reciente ha analizado dos modelos predictivos clínicos que incluían las variables anteriores¹⁵, en forma de escala de puntuación y como algoritmo de decisión clínica. Los autores escogieron puntos de corte y algoritmos que maximizaban la sensibilidad (Se de 96 y 98% para cada modelo), lo que generaba un descenso importante de la especificidad (Es de 36 y 32%). Como consecuencia, los modelos predictivos sólo resultaban útiles en los pacientes de bajo riesgo, que constituían tan sólo una cuarta parte de los pacientes, ya que en ellos la probabilidad de apendicitis ($P_{pre} = 31\%$) se reducía de forma importante (P_{post} negativa de 2 y 4%), mientras que en el resto apenas se modificaba (P_{post} positiva de 40 y 39%). Hubiera sido deseable que los autores ofrecieran información sobre los resultados de los modelos predictivos con otros puntos de corte alternativos, a partir de los cuales pudiéramos calcular CP y probabilidades ajustados a cada nivel de riesgo, y estimar el rendimiento global mediante curvas ROC.

Otro tema de interés lo constituye el cribado universal de hipoacusia neonatal. A pesar de que en nuestro país se está generalizando su uso, una reciente revisión sistemática no ha podido encontrar evidencias sobre la efectividad a largo plazo del cribado universal, en com-

paración con el cribado selectivo a sujetos de riesgo³⁰. Asimismo, aunque distintos programas de cribado universal de hipoacusia han comprobado que aumentan los niños que inician su estudio y tratamiento antes de los 6 meses de edad^{29,32}, queda por demostrar que este objetivo se traduzca en una mejoría, a largo plazo, del nivel educativo y la calidad de vida de los pacientes, que compense de los costes económicos y por falsos positivos de estos programas^{63,64}. Otros trabajos han comparado distintas estrategias de cribado mostrando que un programa de base hospitalaria resulta más económico que uno comunitario²⁷, y que un programa en el que se emplean, de forma escalonada, otoemisiones evocadas y potenciales auditivos automáticos ofrece menos falsos positivos que el uso exclusivo de otoemisiones³⁴.

En cuanto al cribado de displasia de desarrollo de caderas, una reciente revisión sistemática no ha podido encontrar evidencia que apoye la validez o la efectividad del cribado ecográfico en el recién nacido⁴². No cuestionamos la validez del examen ecográfico de cadera, pero sí la utilidad y sobre todo la efectividad del programa universal de cribado. Aunque Roovers et al⁴¹ han observado que el cribado ecográfico universal, en los 3 primeros meses de vida, incrementa los casos de displasia de cadera diagnosticados y tratados, este hallazgo puede deberse a un sobrediagnóstico y consiguiente sobretratamiento, no habiéndose demostrado sus ventajas a largo plazo. Por otra parte, en este estudio⁴¹ se emplearon como patrón de referencia los hallazgos ecográficos a los 6 meses de vida, lo que puede estar originando un sesgo de incorporación que haya distorsionado las estimaciones de validez. En ausencia de nuevos estudios, en el momento actual parece razonable que el cribado ecográfico se reserve a recién nacidos con factores de riesgo⁴⁰.

El patrón de referencia para diagnosticar gastritis ulcus por *Helicobacter pylori* sigue siendo en la actualidad la demostración endoscópica de lesión gástrica y la comprobación del microorganismo en biopsia. Sin embargo, el carácter invasivo de la endoscopia ha motivado el desarrollo de distintas pruebas diagnósticas no invasivas. Nadie discute la utilidad de estas pruebas en el seguimiento

de los pacientes, pero en la infancia, a diferencia del adulto, no se recomienda sustentar únicamente en pruebas no invasivas el diagnóstico y el tratamiento. En el último año, se han publicado dos estudios que valoran la validez de la determinación de antígeno de *H. pylori* en heces^{36,37}. Existen discordancias en los estimadores de validez encontrados, mientras que un estudio, realizado en un solo centro, muestra valores de Se y Es de 97 y 98%³⁷, en otro trabajo multicéntrico sólo alcanzan cifras de 80 y 93%, respectivamente³⁶. Estas diferencias podrían deberse a que el patrón de referencia en el primer trabajo sólo incluía el estudio histológico y el test de la ureasa de la biopsia, mientras que en el segundo también se incluía el cultivo; otras posibles explicaciones son la existencia de heterogeneidad en las muestras y problemas de reproducibilidad de la prueba. Asimismo, en el estudio multicéntrico se comparó la validez del test de antígeno en heces con la del test de la Urea-C¹³ en aire espirado, y se encontró que esta última prueba tenía cifras de Se y Es superiores (96% y 97%). Aunque el test antigénico en heces podría ofrecer ventajas en niños pequeños, donde la recogida de aliento resulta dificultosa, la validez de la prueba en estos niños desciende de forma importante. Por otra parte, debemos señalar que la aplicabilidad en nuestra práctica clínica de los resultados de estos trabajos debe ser cautelosa, ya que los pacientes incluidos en ellos son sin duda una muestra seleccionada de pacientes, con alta probabilidad preprueba de gastritis-ulcus por *H. pylori*. Tendremos, por lo tanto, que adaptar los cálculos a nuestra práctica clínica a la hora de estimar su utilidad diagnóstica.

De especial trascendencia para la práctica pediátrica es la técnica de recogida de orina, ante la sospecha de infección urinaria. A pesar de las limitaciones que presenta la recolección de orina mediante bolsa perineal en el lactante, este método sigue siendo ampliamente utilizado. Es

un hecho reconocido que los urocultivos realizados con esta técnica presentan altos porcentajes de falsos positivos. Trabajos recientes han recordado que también la determinación de parámetros urinarios, como la presencia de leucocitos o nitritos, resulta menos fiable si la orina es recogida con bolsa perineal^{45,46}. En un amplio estudio retrospectivo sobre pacientes menores de 3 meses norteamericanos, el 70% de los urocultivos se realizó mediante cateterismo; éstos presentaron un significativamente menor porcentaje de cultivos contaminados⁴⁶. En nuestro país, el uso de técnicas de recogida estéril, como el cateterismo o la punción vesical, es muy poco habitual. Este hecho constituye, sin duda, una importante área de mejora de la práctica clínica.

Si revisamos el comportamiento de otras pruebas diagnósticas, esquemáticamente recogidas en la tabla II, son pocas las que ofrecen modificaciones relevantes de la probabilidad diagnóstica. Por otra parte, si analizamos los estudios que sí ofrecen cambios relevantes de probabilidad, en muchos de ellos las probabilidades preprueba son muy superiores a las que nos encontraremos en nuestra práctica clínica. Sólo el cálculo de las probabilidades posprueba, ajustadas al escenario clínico en que nos planteemos realizar la prueba, nos va a permitir conocer su rendimiento. Por otra parte, tendremos que valorar si el conocimiento de esas nuevas probabilidades cambia en algo nuestro comportamiento diagnóstico o terapéutico para juzgar la verdadera utilidad clínica de cada prueba diagnóstica.

Puede parecer en un principio complejo todo este juego de probabilidades. Sin embargo, la alternativa a conocer y considerar el verdadero papel de las pruebas diagnósticas es entregarnos a una interpretación simplista del proceso diagnóstico, que no beneficiará a nuestros pacientes.

Bibliografía

1. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med*. 1978;299:926-30.
2. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA*. 1995;274:645-51.
3. Shapiro DE. The interpretation of diagnostic tests. *Stat Methods Med Res*. 1999;8:113-34.
4. Zhou XH. Correcting for verification bias in studies of a diagnostic test's accuracy. *Stat Methods Med Res*. 1998;7:337-53.
5. Ochoa Sangrador C. Valoración crítica de documentos científicos. Aplicabilidad de los resultados de la valoración a nuestra práctica clínica. *Bol Pediatr*. 2002;42:120-30.
6. Ochoa Sangrador C, Orejas G. Epidemiología y metodología científica aplicadas a la pediatría (IV): pruebas diagnósticas. *An Esp Pediatr*. 1999;50:301-14.
7. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA*. 1994;271:389-91.
8. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA*. 1994;271:703-7.
9. Aizpurua Galdeano P, Balaguer Santamaría A. Evaluación clínica retrospectiva del manejo del traumatismo craneal. La radiografía de cráneo parece prescindible si se dispone de tomografía computarizada. *Evid Pediatr*. 2005;1:8.
10. Buñuel Álvarez JC, Ochoa Sangrador C. Dos reglas de decisión clínica resultan útiles para identificar a los niños con dolor abdominal con baja probabilidad de tener apendicitis. *Evid Pediatr*. 2006;2:5.
11. Ochoa Sangrador C, Orejón de Luna G. La cistografía tiene una utilidad limitada como predictora del daño renal en niños menores de un año que presentan un primer episodio de infección urinaria y ecografía renal normal. *Evid Pediatr*. 2005;1:3.
12. Perdikidis Olivieri L, González de Dios J. El cribado ecográfico de displasia de desarrollo de caderas sólo se justifica en recién nacidos con determinados factores de riesgo. *Evid Pediatr*. 2005;1:4.
13. Fefferman NR, Bomsztyk E, Yim AM, Rivera R, Amodio JB, Pinkney LP, et al. Appendicitis in children: low-dose CT with a phantom-based simulation technique-initial observations. *Radiology*. 2005;237:641-6.
14. Rhea JT, Halpern EF, Ptak T, Lawrason JN, Sacknoff R, Novelline RA. The status of appendiceal CT in an urban medical center 5 years after its introduction: experience with 753 patients. *AJR Am J Roentgenol*. 2005;184:1802-8.
15. Kharbanda AB, Taylor GA, Fishman SJ, Bachur RG. A clinical decision rule to identify children at low risk for appendicitis. *Pediatrics*. 2005;116:709-16.
16. York D, Smith A, Phillips JD, von Allmen D. The influence of advanced radiographic imaging on the treatment of pediatric appendicitis. *J Pediatr Surg*. 2005;40:1908-11.
17. Garin EH, Olavarría F, García Nieto V, Valenciano B, Campos A, Young L. Clinical significance of primary vesicoureteral reflux and urinary antibiotic prophylaxis after acute pyelonephritis: a multicenter, randomized, controlled study. *Pediatrics*. 2006;117:626-32.
18. Georgaki-Angelaki H, Kostaridou S, Daikos GL, Kapoyiannis A, Veletzas Z, Michos AG, et al. Long-term follow-up of children with vesicoureteral reflux with and without antibiotic prophylaxis. *Scand J Infect Dis*. 2005;37:842-5.
19. Ismaili K, Hall M, Piepsz A, Wissing KM, Collier F, Schulman C, et al. Primary vesicoureteral reflux detected in neonates with a history of fetal renal pelvis dilatation: a prospective clinical and imaging study. *J Pediatr*. 2006;148:222-7.
20. Moorthy I, Easty M, McHugh K, Ridout D, Biassoni L, Gordon I. The presence of vesicoureteric reflux does not identify a population at risk for renal scarring following a first urinary tract infection. *Arch Dis Child*. 2005;90:733-6.
21. Newman TB. Evidence does not support American Academy of Pediatrics recommendation for routine imaging after a first urinary tract infection. *Pediatrics*. 2005;116:1613-4.
22. Wald ER. Vesicoureteral reflux: the role of antibiotic prophylaxis. *Pediatrics*. 2006;117:919-22.
23. Edmondson JD, Maizels M, Alpert SA, Kirsch AJ, Hanna MK, Weiser AC, et al. Multi-institutional experience with PIC cystography-incidence of occult vesicoureteral reflux in children with febrile urinary tract infections. *Urology*. 2006;67:608-11.
24. Leroy S, Adamsbaum C, Marc E, Moulin F, Raymond J, Gendrel D, et al. Procalcitonin as a predictor of vesicoureteral reflux in children with a first febrile urinary tract infection. *Pediatrics*. 2005;115:e706-9.
25. Galanakis E, Bitsori M, Dimitriou H, Giannakopoulou C, Karakavitsas NS, Kalmanti M. Urine interleukin-8 as a marker of vesicoureteral reflux in infants. *Pediatrics*. 2006;117:e863-7.
26. Leroy S, Marc E, Adamsbaum C, Gendrel D, Breart G, Chalumeau M. Prediction of vesicoureteral reflux after a first febrile urinary tract infection in children: validation of a clinical decision rule. *Arch Dis Child*. 2006;91:241-4.
27. Grill E, Uus K, Hessel F, Davies L, Taylor RS, Wasem J, et al. Neonatal hearing screening: modelling cost and effectiveness of hospital and community based screening. *BMC Health Serv Res*. 2006;6:14.
28. Johnson JL, White KR, Widen JE, Gravel JS, James M, Kennalley T, et al. A multicenter evaluation of how many infants with permanent hearing loss pass a two-stage otoacoustic emissions/automated auditory brainstem response newborn hearing screening protocol. *Pediatrics*. 2005;116:663-72.
29. Kennedy C, McCann D, Campbell MJ, Kimm L, Thornton R. Universal newborn screening for permanent childhood hearing loss.

- ring impairment: an 8-year follow-up of a controlled trial. *Lancet*. 2005;366:660-2.
30. Puig T, Municio A, Medà C. Universal neonatal hearing screening versus selective screening as part of the management of childhood deafness. *The Cochrane Database of Systematic Reviews* 2005, Issue 2. Art. No.: CD003731. DOI: 10.1002/14651858.CD003731.
 31. Schroeder L, Petrou S, Kennedy C, McCann D, Law C, Watkin PM, et al. The economic costs of congenital bilateral permanent childhood hearing impairment. *Pediatrics*. 2006;117:1101-12.
 32. Uus K, Bamford J. Effectiveness of population-based newborn hearing screening in England: ages of interventions and profile of cases. *Pediatrics*. 2006;117:e887-93.
 33. Uus K, Bamford J, Taylor R. An analysis of the costs of implementing the National Newborn Hearing Screening Programme in England. *J Med Screen*. 2006;13:14-9.
 34. Lin HC, Shu MT, Lee KS, Ho GM, Fu TY, Bruna S, et al. Comparison of hearing screening programs between one step with transient evoked otoacoustic emissions (TEOAE) and two steps with TEOAE and automated auditory brainstem response. *Laryngoscope*. 2005;115:1957-62.
 35. Harris P, Pérez-Pérez G, Zylberberg A, Rollan A, Serrano C, Riera F, et al. Relevance of adjusted cut-off values in commercial serological immunoassays for *Helicobacter pylori* infection in children. *Dig Dis Sci*. 2005;50:2103-9.
 36. Megraud F. Comparison of non-invasive tests to detect *Helicobacter pylori* infection in children and adolescents: results of a multicenter European study. *J Pediatr*. 2005;146:198-203.
 37. Sabbi T, De Angelis P, Colistro F, Dall'Oglio L, Di Abriola GF, Castro M. Efficacy of noninvasive tests in the diagnosis of *Helicobacter pylori* infection in pediatric patients. *Arch Pediatr Adolesc Med*. 2005;159:238-41.
 38. Joram N, Boscher C, Denizot S, Loubersac V, Winer N, Roze JC, et al. Umbilical cord blood procalcitonin and C reactive protein concentrations as markers for early diagnosis of very early onset neonatal infection. *Arch Dis Child Fetal Neonatal Ed*. 2006;91:F65-6.
 39. Garges HP, Moody MA, Cotten CM, Smith PB, Tiffany KF, Lenfestey R, et al. Neonatal meningitis: what is the correlation among cerebrospinal fluid cultures, blood cultures, and cerebrospinal fluid parameters? *Pediatrics*. 2006;117:1094-100.
 40. Lowry CA, Donoghue VB, Murphy JF. Auditing hip ultrasound screening of infants at increased risk of developmental dysplasia of the hip. *Arch Dis Child*. 2005;90:579-81.
 41. Roovers EA, Boere-Boonekamp MM, Castelein RM, Zielhuis GA, Kerkhoff TH. Effectiveness of ultrasound screening for developmental dysplasia of the hip. *Arch Dis Child Fetal Neonatal Ed*. 2005;90:F25-30.
 42. Woolacott NF, Puhan MA, Steurer J, Kleijnen J. Ultrasonography in screening for developmental dysplasia of the hip in newborns: systematic review. *BMJ*. 2005;330:1413.
 43. Kalach N, Soulaïnes P, de Boissieu D, Dupont C. A pilot study of the usefulness and safety of a ready-to-use atopy patch test (Diallertest) versus a comparator (Finn Chamber) during cow's milk allergy in children. *J Allergy Clin Immunol*. 2005;116:1321-6.
 44. Kappler M, Krauss-Etschmann S, Diehl V, Zeilhofer H, Koletzko S. Detection of secretory IgA antibodies against gliadin and human tissue transglutaminase in stool to screen for coeliac disease in children: validation study. *BMJ*. 2006;332:213-4.
 45. McGillivray D, Mok E, Mulrooney E, Kramer MS. A head-to-head comparison: "clean-void" bag versus catheter urinalysis in the diagnosis of urinary tract infection in young children. *J Pediatr*. 2005;147:451-6.
 46. Schroeder AR, Newman TB, Wasserman RC, Finch SA, Pantell RH. Choice of urine collection methods for the diagnosis of urinary tract infection in young, febrile infants. *Arch Pediatr Adolesc Med*. 2005;159:915-22.
 47. Huynh SC, Ojaimi E, Robaei D, Rose K, Mitchell P. Accuracy of the Lang II stereotest in screening for binocular disorders in 6-year-old children. *Am J Ophthalmol*. 2005;140:1130-2.
 48. Nassif DS, Piskun NV, Hunter DG. The Pediatric Vision Screener III: detection of strabismus in children. *Arch Ophthalmol*. 2006;124:509-13.
 49. Reed MJ, Browning JG, Wilkinson AG, Beattie T. Can we abolish skull X rays for head injury? *Arch Dis Child*. 2005;90:859-64.
 50. Rimoin AW, Hamza HS, Vince A, Kumar R, Walker CF, Chitale RA, et al. Evaluation of the WHO clinical decision rule for streptococcal pharyngitis. *Arch Dis Child*. 2005;90:1066-70.
 51. Sarles J, Berthezene P, Le Louarn C, Somma C, Perini JM, Catheline M, et al. Combining immunoreactive trypsinogen and pancreatitis-associated protein assays, a method of newborn screening for cystic fibrosis that avoids DNA analysis. *J Pediatr*. 2005;147:302-5.
 52. Ferrara G, Losi M, D'Amico R, Roversi P, Piro R, Meacci M, et al. Use in routine clinical practice of two commercial blood tests for diagnosis of infection with *Mycobacterium tuberculosis*: a prospective study. *Lancet*. 2006;367:1328-34.
 53. Shams H, Weis SE, Klucar P, Lalvani A, Moonan PK, Pogoda JM, et al. Enzyme-linked immunospot and tuberculin skin testing to detect latent tuberculosis infection. *Am J Respir Crit Care Med*. 2005;172:1161-8.
 54. Swingler GH, Du Toit G, Andronikou S, Van der Merwe L, Zar HJ. Diagnostic accuracy of chest radiography in detecting mediastinal lymphadenopathy in suspected pulmonary tuberculosis. *Arch Dis Child*. 2005;90:1153-6.
 55. Ullrich C, Wu A, Armsby C, Rieber S, Wingerter S, Brugnara C, et al. Screening healthy infants for iron deficiency using reticulocyte hemoglobin content. *JAMA*. 2005;294:924-30.
 56. Vyas H, Hejlik J, Ackerman MJ. Epinephrine QT stress testing in the evaluation of congenital long-QT syndrome: diagnostic accuracy of the paradoxical QT response. *Circulation*. 2006;113:1385-92.
 57. Hosoya M, Kawasaki Y, Katayose M, Sakuma H, Watanabe M,

- Igarashi E, et al. Prognostic predictive values of serum cytochrome c, cytokines, and other laboratory measurements in acute encephalopathy with multiple organ failure. *Arch Dis Child*. 2006;91:469-72.
58. Shah SS, Shofer FS, Seidel JS, Baren JM. Significance of extreme leukocytosis in the evaluation of febrile children. *Pediatr Infect Dis J*. 2005;24:627-30.
59. Korff CM, Nordli Jr. DR. Paroxysmal events in infants: persistent eye closure makes seizures unlikely. *Pediatrics*. 2005; 116:e485-6.
60. Cianfarani S, Liguori A, Boemi S, Maghnie M, Iughetti L, Wasniewska M, et al. Inaccuracy of insulin-like growth factor (IGF) binding protein (IGFBP)-3 assessment in the diagnosis of growth hormone (GH) deficiency from childhood to young adulthood: association to low GH dependency of IGF-II and presence of circulating IGFBP-3 18-kilodalton fragment. *J Clin Endocrinol Metab*. 2005;90:6028-34.
61. Fernández JM, Málaga S. ¿Es posible cambiar el paradigma reflujo-céntrico? *Evid Pediatr*. 2006;2:17.
62. Sargent MA. What is the normal prevalence of vesicoureteral reflux? *Pediatr Radiol*. 2000;30:587-93.
63. González de Dios J, Mollar Maseres J. Cribado universal de hipoacusia neonatal: evaluación de la prueba frente a evaluación del programa. *Acta Otorrinolaringol Esp*. 2005;56:331-4.
64. González de Dios J, Mollar Maseres J, Rebagliato Russo M. Evaluación del programa de detección precoz universal de la hipoacusia en el recién nacido. *An Pediatr (Barc)*. 2005;63: 230-7.