

J. González de Dios

An Esp Pediatr 1999;51:461-468.

Introducción

El área de investigación en pruebas diagnósticas es, junto con la de innovaciones terapéuticas, una de las áreas con mayor y más rápido desarrollo en la medicina actual. Es posible que este rápido desarrollo de nuevos procedimientos diagnósticos no haya tenido un paralelismo en el desarrollo de las más adecuadas estrategias de evaluación. Y en este sentido sí encontramos diferencias destacables con el proceso previo a la introducción de un nuevo fármaco, sometido a estrictos requisitos legales y a una evaluación formalizada, que generalmente toma la forma del ensayo clínico aleatorizado^(1,2).

Los rápidos avances en los métodos diagnósticos implican la necesidad de que los profesionales dispongan de la información correcta sobre sus características y la aplicabilidad en su ámbito de trabajo. Así pues, la investigación sobre pruebas diagnósticas tiene como objetivos el de estimar la capacidad discriminativa de una prueba diagnóstica entre enfermos y no enfermos (sensibilidad-especificidad), determinar el rendimiento de la misma (valores predictivos) o evaluar la utilidad y satisfacción de un procedimiento diagnóstico.⁽³⁻⁵⁾

Se han publicado numerosas recomendaciones sobre la valoración de pruebas diagnósticas, algunas desde una perspectiva tan actual como la que ofrece la medicina basada en la evidencia (o en pruebas)⁽⁶⁻⁹⁾. A pesar de estas recomendaciones es evidente la limitación metodológica en esta área de la investigación, lo cual se ha puesto de manifiesto incluso en las mejores revistas biomédicas internacionales⁽¹⁰⁾. Esta carencia de rigurosidad en la investigación sobre pruebas diagnósticas, puede conllevar a que la aplicabilidad de la medicina basada en la evidencia pueda verse limitada en el campo del diagnóstico⁽¹¹⁾. La utilidad de una prueba diagnóstica depende de su capacidad de producir los mismos resultados cada vez que se aplica en similares condiciones (fiabilidad) y de que sus mediciones reflejen exactamente el fenómeno que se intenta medir (validez o exactitud).

En España se han tratado los aspectos más teóricos de la investigación sobre pruebas diagnósticas^(4,5,12), pero sólo conocemos dos estudios sobre la calidad de la metodología empleada, aplicados sobre revistas de medicina general^(3,13). No conocemos

Valoración de la metodología en los estudios sobre pruebas diagnósticas en "Anales Españoles de Pediatría"

ningún estudio sobre la valoración de la metodología de los estudios sobre pruebas diagnósticas en el campo de la pediatría.

La revista *Anales Españoles de Pediatría* (*An Esp Pediatr*) se ha consolidado como la revista pediátrica de mayor impacto en lengua castellana, y ocupa un lugar destacado dentro de las revistas médicas españolas en relación con los indicadores bibliométricos de calidad científica: factor de impacto, nº de trabajos, nº de revistas citadoras, índice de visibilidad...⁽¹⁴⁾. En *An Esp Pediatr* se publican aproximadamente la cuarta parte de todos los documentos de las revistas pediátricas españolas, y actúa como publicación oficial de la Asociación Española de Pediatría, siendo la revista pediátrica con mayor impacto en lengua castellana^(15,16). Así pues, *An Esp Pediatr* contiene una buena representación de las mejores investigaciones sobre pruebas diagnósticas en pediatría realizadas en España.

El **objetivo** de nuestro trabajo es realizar una evaluación de la metodología empleada en los artículos originales sobre pruebas diagnósticas publicados en *An Esp Pediatr* durante un período de 5 años (años 1994 a 1998) con la finalidad de describir su adecuación a estándares metodológicos aceptados y empleados anteriormente^(3,10,13).

Material y métodos

* Selección de los artículos

Se revisaron los trabajos sobre pruebas diagnósticas publicados como artículos originales en *An Esp Pediatr* desde 1994 a 1998. Tras la lectura de los artículos se seleccionaron aquellos trabajos que valoran pruebas diagnósticas en términos de sensibilidad, especificidad o sustitutos adecuados.

* Guía de calidad metodológica

Realizamos las modificaciones oportunas para abreviar la ficha de calidad metodológica utilizada previamente en nuestro país por I. Hernández y cols.^(3,13) que está basada en la combinación de criterios propios y de otros autores^(7,17-20). Se divide en 6 apartados con un número variable de ítems, tal como se especifica en el Anexo.

A- **Objetivos del estudio**: Este apartado recoge la especificación del objetivo del estudio y su justificación. La valoración de una prueba diagnóstica está justificada si se trata de una prueba nueva, de una en uso pero con una nueva aplicación o si hay controversias en su validez.

B- **Patrón de referencia**: La prueba de referencia (PR) debe

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

Correspondencia: Dr. Javier González de Dios. C/ Prof. Manuel Sala, 6. 3º A. 03003 Alicante

estar reconocida como tal por la comunidad científica, se debe aplicar a toda la serie de casos estudiados y no tiene que incorporar información procedente de la prueba diagnóstica que se evalúa (sesgo de incorporación). Si se trata de una PR compleja (mezcla de criterios clínicos y/o batería de pruebas), debe estandarizarse y aplicarse sistemáticamente describiendo todo ello de forma explícita en el trabajo.

C- Método de realización de la prueba diagnóstica: Descripción adecuada del procedimiento de realización de la prueba diagnóstica, el estudio de su reproductibilidad y la definición del término normal y/o anormal, así como la justificación del criterio de normalidad elegido.

D- Diseño del estudio: En este apartado se recogen el tipo y la direccionalidad del estudio (importante para las conclusiones del mismo), el ámbito de realización y la población de estudio, los criterios de selección y exclusión, el espectro de los sujetos, y la presencia de los sesgos de secuencia diagnóstica y de revisión.

El tipo de diseño se basa en la clasificación de Feinstein⁽²¹⁾, que los agrupa en cuatro categorías: I, comparación de enfermos y sanos; II y III, comparación ampliando el espectro a distinta gravedad de la enfermedad y a otras enfermedades de diagnóstico diferencial; y IV, referido a la utilización de la prueba en la práctica clínica habitual; I. Hernández y cols⁽¹³⁾ también incluyen una categoría 0, en la que se encontrarían los estudios que sólo incluyen a los sujetos enfermos.

Los sesgos de secuencia y de revisión se definen en base a los criterios de Reid y cols⁽¹⁰⁾. Conviene evitar ambos sesgos en la valoración de las pruebas diagnósticas:

-Evitar el sesgo de secuencia (verificación diagnóstica o *verification bias*). En los estudios de cohortes se aceptaba el criterio si a todos los sujetos se les practicaba la prueba diagnóstica y el patrón de referencia, este último directamente o mediante seguimiento clínico. En los estudios de casos y controles en los que la prueba diagnóstica precedía a la PR, se cumplía el estándar si la verificación de la enfermedad se realizaba en una serie consecutiva de pacientes con independencia del resultado de la prueba diagnóstica. En aquellos estudios de casos y controles en los que la PR precedía a la prueba diagnóstica, el crédito se obtenía cuando los resultados de ésta eran estratificados de acuerdo con los factores clínicos que potencialmente condicionaban la indicación de la PR.

-Evitar el sesgo de revisión (valoración ciega o *review bias*). Para estudios de cohortes prospectivos en los que el paciente recibía primeramente la prueba diagnóstica, el estándar era aceptado si la PR se evaluaba independientemente, es decir, sin conocer el resultado de la prueba diagnóstica. En estudios prospectivos donde la PR precedía a la prueba diagnóstica y en los estudios de casos y controles, se cumplía el estándar si figuraba una afirmación de la independencia de interpretación de los datos.

E- Resultados: Este apartado recoge el tipo de índices empleados (sensibilidad y/o especificidad, cocientes de verosimilitud o sustitutos adecuados), la conveniencia de los cálculos de

Tabla I Estudio sobre pruebas diagnósticas en relación con el número total de artículos originales

Año	Nº originales	Nº estudios sobre pruebas diagnósticas (%)*
1994	106	4 (3,8%)
1995	83	3 (3,6%)
1996	143	8 (5,6%)
1997	130	5 (3,8%)
1998	114	6 (5,3%)

*El porcentaje expresa el número de estudios sobre pruebas diagnósticas sobre el número total de artículos originales de *An Esp Pediatr*.

los valores predictivos y de las curvas ROC (*receiver operating characteristics*) y el empleo de índices conjuntos en el caso de estudiar varias pruebas diagnósticas. Se examinó si los autores calcularon la precisión de los índices de exactitud (error estándar o intervalo de confianza) y si comunicaban los resultados por estratos clínicos relevantes, es decir, comprobaban si la exactitud de la prueba diagnóstica variaba según el grupo de pacientes en los que se aplicaba.

F- Conclusiones: En este apartado se pretende recoger la valoración global sobre la adecuación de las conclusiones de los autores según el tipo de estudio y la metodología empleada, así como la interpretación apropiada o no de los valores predictivos.

* Metodología del estudio

La ficha de calidad metodológica fue aplicada a la totalidad de los artículos que cumplieron los criterios de selección. Se ha considerado el número de sujetos (o, en su defecto, muestras) incluidos en cada uno de los trabajos, calculándose la distribución de frecuencias y la media. También se ha evaluado el tipo de prueba diagnóstica y la especialidad clínica de aplicación preferente de dicha prueba. Para este último particular se consideró también el servicio de procedencia de la mayoría de los autores del estudio.

La creación, la gestión de la base de datos y los análisis estadísticos de los resultados se efectuaron con el programa EPIINFO.

Resultados

Durante el período de estudio (1994-98) se revisaron el conjunto de los 576 artículos originales de *An Esp Pediatr*, de los cuales 26 (4,5%) cumplían los criterios de selección de estudios sobre pruebas diagnósticas. La lista de estudios evaluados figura en las referencias bibliográficas como "artículos revisados". En la tabla I se pormenoriza la distribución de frecuencias por cada año de estudio.

La mediana del número de sujetos/muestras era de 122, con un intervalo muy amplio (39-9.000) y una distribución de frecuencias que se refleja en la tabla II.

Tabla II Distribución de frecuencias del número de sujetos/muestras evaluadas en los 26 artículos seleccionados

Nº sujetos/muestras	Número (%)
39 - 99	8 (30,8%)
100 - 199	8 (30,8%)
200 - 399	4 (15,4%)
400 - 999	1 (3,8%)
≥ 1.000	5 (19,2%)

Las subespecialidades pediátricas en que con mayor frecuencia se realizaron estudios sobre pruebas diagnósticas son neonatología (7 artículos), infecciosos (5) y neumología (4) (Tabla III).

Los procedimientos de laboratorio, tales como bioquímica, inmunología, serología o microbiología, fueron los evaluados más frecuentemente (65,4%) (Tabla IV).

Al aplicar la ficha de calidad metodológica reseñada en el Anexo para la valoración de la metodología aplicada a los estudios sobre pruebas diagnósticas, se obtienen los siguientes resultados:

A- **Objetivos del estudio:** En la mayoría de los artículos (96%) se especifican los objetivos del estudio y en el 84% se justifica adecuadamente la necesidad de evaluar la prueba diagnóstica.

B- **Patrón de referencia:** (Tabla V). En el 73% de los trabajos se especifica claramente la PR, pero sólo el 27% indican si la PR está reconocida como tal por la comunidad científica. En al menos un 42% de los estudios existe posibilidad de sesgo de incorporación, dado que la PR incorpora información de la prueba diagnóstica que se evalúa. Encontramos que un 27% de los estudios estandariza la PR, debido a su mayor complejidad.

C- **Método de realización de la prueba diagnóstica:** La prueba se describe adecuadamente en la mayoría de los trabajos (92%), pero en sólo dos estudios dicha prueba se describe con referencias bibliográficas. Se define el término normal o anormal en la mitad de los estudios, pero sólo se evalúa la reproducibilidad en el 27%.

D- **Diseño del estudio:** (Tabla VI). Predominan los estudios de la categoría II, III y IV de la clasificación de Feinstein. Los estudios de carácter prospectivo se registran con una frecuencia dos veces mayor a los retrospectivos, si bien en un 43% de los trabajos es incierta la direccionalidad del estudio. El ámbito en el que se realiza la investigación con la descripción suficiente del origen de la población a estudio se recoge en el 77% de los estudios. Se especifican correctamente los criterios de selección y de exclusión en el 50 y el 19%, respectivamente. Se describe de forma óptima la muestra de sujetos/especímenes en el 43% de los trabajos. Encontramos la posibilidad de haber evitado el sesgo de secuencia en 16 trabajos (62%) y el de revisión en 17 (65%). Respecto a la secuencia de las pruebas, encontramos que la prueba diagnóstica generalmente antecede a la PR (34%) o se realiza de forma simultánea a ésta (43%).

Tabla III Subespecialidades pediátricas representadas en los 26 estudios de evaluación de pruebas diagnósticas revisados

Subespecialidad pediátrica	Nº artículos
Neonatología	7
Infecciosos	5
Neumología	4
Nefrología	3
Cirugía	3
Metabolismo-nutrición	2
Gastroenterología	1
ORL	1

E- **Resultados:** (Tabla VII). El 92% de los estudios expresan los resultados como sensibilidad y/o especificidad; los valores predictivos se estimaron en el 69%, y tan sólo el 23% presentaron curvas ROC. Destaca que en tan sólo uno de los 26 estudios (4%) determinó la precisión estadística de las estimaciones. Casi la mitad de los artículos mostraban los índices de exactitud por estratos clínicos relevantes.

F- **Conclusiones:** En el aspecto formal, cabe destacar que en el 77% de los estudios las conclusiones eran correctas en general, en el 19% presentaban alguna incorrección, y sólo un estudio no presentaba conclusiones.

Discusión

La metodología empleada en nuestro estudio se fundamenta en estudios previos realizados en nuestro país sobre pruebas diagnósticas en revistas de medicina interna^(3,13). No tenemos conocimiento de estudios similares sobre artículos pediátricos en nuestro entorno, por lo que basamos nuestro estudio en la revista *An Esp Pediatr*, que actúa como publicación oficial de la Asociación Española de Pediatría y se ha consolidado como una de las revistas pediátricas de mayor impacto en castellano^(15,16). Los datos anteriores posibilitan que, aunque los resultados obtenidos en el presente estudio sólo sean válidos para los artículos publicados en la revista *An Esp Pediatr*, pueda realizarse una cierta extrapolación con la producción científica pediátrica en España, siendo conscientes de las limitaciones y sesgos, dado que no se incluyen los trabajos publicados en revistas extranjeras y que pueden ser de mayor calidad. Sin embargo, no es nuestra intención realizar una estimación general de la calidad científica de todos los estudios de este tipo en pediatría, sino detectar los problemas metodológicos más frecuentes en los trabajos publicados en esta revista con un nivel aceptable de calidad científica.

Sobre las características generales de los artículos revisados hay algunos detalles a destacar. Sólo un 4,5% de los artículos originales de *An Es Pediatr* son estudios sobre pruebas diagnósticas, con mínimas diferencias entre los distintos años de estudio (Tabla I). Dos terceras partes de los estudios se realizaron con muestras menores de 200 sujetos/especímenes (Tabla

Tabla IV Categorías de las pruebas diagnósticas evaluadas

	Número (%)
Laboratorio	17 (65,4%)
Criterios clínicos	4 (15,4%)
Pruebas funcionales	4 (15,4%)
Diagnóstico por imagen	1 (3,8%)

II); sin embargo, sólo dos estudios (7,7%) incluían menos de 50 sujetos/ especímenes, que es el tamaño de muestral mínimo aconsejable para evitar estimaciones demasiado imprecisas.

Pocas subespecialidades pediátricas están representadas en los estudios sobre pruebas diagnósticas (Tabla III), y casi la mitad de los artículos se incluyen en las áreas de neonatología e infecciosos, que curiosamente se corresponden con las áreas temáticas de mayor productividad científica en nuestro entorno pediátrico, tal como se ha analizado en estudios bibliométricos previos^(15, 22). La mayoría de las pruebas diagnósticas evaluadas son de laboratorio (Tabla IV), lo que es común en otros estudios similares^(3,13); destacar la escasa o nula representatividad de otras áreas de estudio (diagnóstico por imagen, anatomía patológica).

Nuestros resultados muestran, en términos generales, frecuentes problemas metodológicos en los estudios de valoración de pruebas diagnósticas, algunos de los cuales comprometen la interpretación de los resultados y la utilidad final de la información producida.

La evaluación de la exactitud de una determinada prueba diagnóstica se basa, generalmente, en la comparación de los resultados de ésta con una PR o "patrón oro" (*gold standard*). La falta de especificación de la PR, las dudas sobre la aceptación por la comunidad científica o su aplicación de forma heterogénea compromete la calidad científica de los resultados. Todos ellos son defectos hallados con excesiva frecuencia en nuestra serie (Tabla V). Además, en el 42% de los estudios se valoraban pruebas diagnósticas cuyos resultados se empleaban como criterios para el diagnóstico final, es decir, pruebas diagnósticas que formaban parte de la PR, dando lugar con ello a una sobreestimación de los índices de exactitud (sesgo de incorporación). Es importante tener en cuenta que una PR puede ser imperfecta o, en ocasiones, puede resultar imposible establecer una PR al estudiar determinada prueba diagnóstica; en estos casos se tratará al menos de valorar la fiabilidad de la prueba diagnóstica y su concordancia con otras pruebas alternativas.

Como aspecto positivo hay que destacar que en la mayoría de los trabajos revisados (92%) se detalla adecuadamente el método de realización de la prueba diagnóstica, si bien sería deseable que se avalase esta descripción con referencias bibliográficas. Cuando la interpretación de la prueba diagnóstica en cuestión tiene un importante componente de subjetividad, es recomendable evaluar, si es posible, la variabilidad inter e intraobservador: la reproductibilidad de la prueba sólo constaba en el 27% de los artículos revisados.

Tabla V Aspectos relativos al patrón de referencia de los artículos seleccionados

	Número	(%)
Especificado	19	(73%)
Indicada su aceptación	7	(27%)
Incorporación de la prueba a valorar:		
Sí	11	(42%)
No especificado	5	(19%)
Realizado en toda la serie:		
Sí	21	(81%)
No especificado	2	(8%)
Estandarizado	7	(27%)

Para valorar la posible aplicabilidad de una prueba diagnóstica en un ámbito de trabajo determinado es necesario considerar si lo observado en el estudio es extrapolable a nuestra práctica clínica, de ahí lo importante de los aspectos relativos a la descripción del estudio. Aunque la descripción del ámbito del estudio se realiza total o parcialmente en la mayoría de los artículos, llaman la atención las deficiencias observadas en la descripción de los criterios de selección y exclusión, así como la descripción del espectro de sujetos (Tabla VI), lo que limita de alguna forma la información que aportan los trabajos y las posibilidades que tiene el lector de aplicar los resultados a sus propios pacientes.

Hay sesgos que pueden alterar la valoración de una prueba diagnóstica como el de secuencia diagnóstica y el de revisión⁽²³⁾. El sesgo de secuencia diagnóstica aparece cuando no todos los sujetos tienen la misma probabilidad de que se confirme la enfermedad, hecho que suele ocurrir cuando la PR es agresiva y supone riesgos para el paciente. Este sesgo ocurrirá cuando la probabilidad de que se realice la PR sea menor entre los sujetos con la prueba diagnóstica negativa y, por lo tanto, sea menos probable que éstos entren en el estudio; este sesgo producirá sobreestimaciones de la sensibilidad e infraestimaciones de la especificidad. En trabajos previos este sesgo se presenta entre el 26 y 36% de los trabajos^(3,24,25) y en nuestro artículo aparece en porcentajes algo superiores (seguro en el 19% y sospechoso en el 43%). El sesgo de revisión aparece cuando la prueba a valorar y la PR no se evalúan independientemente, es decir, cuando la persona que evalúa la PR conoce el resultado de la prueba a valorar, y/o viceversa^(26,27). Este sesgo estaba presente en el 65% de nuestros artículos, cifra similar a la detectada en estudios similares⁽³⁾.

El proceso diagnóstico es, en esencia, un cálculo de probabilidades. La principal cualidad de una prueba diagnóstica es su validez o exactitud. Clásicamente la exactitud de una prueba diagnóstica se ha evaluado en función de dos características: la sensibilidad y la especificidad, cuyos resultados dependen del valor que se elige como el "punto de corte" para diferenciar entre normal y anormal. Hay una relación recíproca entre sensibilidad y especificidad, cuyos resultados son independientes de la prevalencia de la enfermedad, a diferencia de los valores pre-

Tabla VI Aspectos relativos al diseño del estudio de los artículos seleccionados

	Número	(%)
Fase del estudio		
0	3	(12%)
I	5	(19%)
II-III	10	(38%)
IV	8	(31%)
Direccionalidad:		
Prospectiva	10	(38%)
Retrospectiva	5	(19%)
Incierta	11	(43%)
Descripción del ámbito de estudio:		
Sí	20	(77%)
Parcialmente	5	(19%)
Criterios de selección:		
Sí	13	(50%)
Parcialmente	10	(38%)
Criterios de exclusión:		
Sí	5	(19%)
Parcialmente	3	(12%)
Descripción del espectro:		
Sí	11	(43%)
Parcialmente	3	(12%)
Sesgo de secuencia:		
Seguro	5	(19%)
Sospechoso	11	(43%)
Improbable	10	(38%)
Comparación ciega:		
No	17	(65%)
No se especifica	8	(31%)
Secuencia:		
Primero prueba de referencia	2	(8%)
Primero prueba a valorar	9	(34%)
Simultánea	11	(43%)
Incierta	4	(15%)

dictivos: así, cuanto mayor sea la prevalencia de la enfermedad mayor será el valor predictivo positivo y menor el valor predictivo negativo^(4,5). Una forma más global de conocer la calidad de la prueba es mediante el uso de curvas ROC, que es un gráfico en el que se observan todos los pares sensibilidad/especificidad resultantes de la variación continua de los puntos de corte; las curvas ROC permiten explorar la capacidad diagnóstica de la prueba en sus distintos valores, de manera que podamos conocer su validez global y seleccionar el punto o puntos de corte más adecuados⁽¹²⁾. En nuestros artículos seleccionados los resultados se expresaron en un 92% como sensibilidad y/o especificidad, un 69% como valor predictivo y un 23% como pruebas ROC.

Tabla VII Aspectos relativos a la presentación de los resultados de los artículos seleccionados

	Número	(%)
Sensibilidad y/o especificidad	24	(92%)
Curvas ROC	6	(23%)
Índices para pruebas conjuntas	6	(23%)
Valor predictivo	18	(69%)
Medición error aleatorio	1	(4%)
Resultados por estratos	11	(43%)

Cabe destacar en la presentación de los resultados de la investigación en pruebas diagnósticas la importancia de las precisiones estadísticas de las estimaciones⁽¹⁰⁾ y la presentación de los resultados por estratos clínicos relevantes⁽²⁸⁾. En los últimos años se está teniendo más en cuenta la importancia de la precisión estadística de las estimaciones (intervalo de confianza o error estándar)^(29,30) pero sólo lo detectamos en uno de nuestros artículos revisados (4%). Dado que los descriptores de validez de las pruebas diagnósticas son estimaciones puntuales de los verdaderos descriptores poblacionales, están sujetos a variabilidad aleatoria y, por lo tanto, deben proporcionarse con la precisión estadística de las estimaciones; la precisión estadística nos sirve para valorar y diferenciar entre lo estadísticamente significativo y lo clínicamente importante⁽³¹⁾. También es necesaria la presentación de los resultados por estratos con importancia clínica, ya que permite hacer una interpretación ajustada a la práctica clínica diaria. En este sentido, un 43% de nuestros trabajos presentaban las estimaciones por subgrupos.

Como **conclusión** de nuestro estudio hay que destacar la escasa frecuencia de estudios sobre pruebas diagnósticas (4,5 del total de artículos originales en *An Esp Pediatr*), centrado en pocas subespecialidades pediátricas (principalmente neonatología e infecciosos), y en el que se constatan problemas metodológicos similares a los detectados en estudios previos realizados en otras áreas de la medicina en nuestro país^(3,13). Los principales puntos a mejorar son: mejor descripción de los criterios de inclusión/exclusión, evitar los sesgos más comunes en pruebas diagnósticas (sesgos de incorporación, de secuencia y de revisión), y presentar los resultados con mayor precisión estadística (intervalo de confianza o error estándar).

Una evaluación rigurosa de la calidad de las pruebas diagnósticas antes de generalizar su uso permite eliminar pruebas inútiles o de poco valor, mejorar la calidad de la información disponible sobre la aplicabilidad de la prueba diagnóstica, reducir los costes sanitarios y mejorar la asistencia al paciente^(3,10). Muchos de los problemas detectados en nuestro estudio no invalidan los resultados de los trabajos, si acaso limitan el uso de la información aportada y la calidad de su evidencia científica⁽³²⁾ y son problemas que se podrían haber solucionado en el proce-

so de revisión. Confiamos en que la difusión del Anexo (ficha de calidad metodológica para la evaluación de los estudios sobre pruebas diagnósticas) sea de utilidad como lista-guía para los autores y revisores de este tipo de estudios en pediatría, lo que sin duda contribuirá a mejorar su calidad científica.

Addendum: Tras la elaboración de este artículo, se ha publicado en la sección de Educación continuada de *An Esp Pediatr* una magnífica revisión sobre pruebas diagnósticas⁽³³⁾ en el que el lector interesado podrá profundizar en los aspectos teóricos de este tipo de estudios.

Anexo Ficha de calidad metodológica para la valoración de estudios sobre pruebas diagnósticas

A. Objetivos del estudio		D3: Descripción ámbito de realización (fuente de la población a estudio):	
A1: ¿Se especifican?	Sí No		Sí No
A2: ¿Se justifica la valoración de la prueba?	Sí No	D4: ¿Se explican los criterios de selección?	Sí No Parcialmente
B. Patrón de referencia		D5: ¿Se explican los criterios de exclusión?	
B1: ¿Se especifica?	Sí No		Sí No Parcialmente
B2: ¿Se indica si es aceptado?	Sí No	D6: Descripción del espectro de los sujetos:	Sí No Parcialmente
B3: ¿Incorpora pruebas a valorar?	Sí No No se especifica	E. Resultados	
B4: ¿Realizado en toda la serie?	Sí No No se especifica	E1: Se expresa sensibilidad, especificidad o sustitutos adecuados:	Sí No
B5: ¿Se estandariza la prueba de referencia?	Sí No	E2: ¿Curvas ROC o cocientes de verosimilitud?	Sí No
C. Método de realización de la prueba diagnóstica		E3: Índices para pruebas conjuntas:	Sí No
C1: Suficientemente descrito:	Sí Sí con referencias No	E4: Valor predictivo:	Sí No
C2: Reproducibilidad:	Sí No	E5: Medición de error aleatorio:	Sí No
C3: Definición del término normal:	Sí No	E6: Resultados por estratos:	Sí No
D. Diseño del estudio		F. Conclusiones	
D1: Fase del estudio diagnóstico:	0 I II-III IV	F1: Conclusiones correctas:	Correctas en general Alguna incorrección Incorrectas en general Sin conclusiones
D2: Direccionalidad:	Prospectivo Retrospectivo Ambispectivo Incierto		

Bibliografía

- Galende I, Sacristán JA, Soto J. Cómo mejorar la calidad de los ensayos clínicos. *Med Clin (Barc)* 1994; **102**:465-470.
- González De Dios J. Las mejores evidencias (o pruebas) científicas (I): el ensayo clínico. *Rev Esp Pediatr* 1999; **55**:157-165.
- Ramos JM, Hernández I. Investigación sobre pruebas diagnósticas en Medicina Clínica. Valoración de la metodología. *Med Clin (Barc)* 1998; **111**:129-134.
- Pozo F. La eficacia de las pruebas diagnósticas (I). *Med Clin (Barc)* 1988; **90**:779-785.
- Pozo F. La eficacia de las pruebas diagnósticas (y II). *Med Clin (Barc)* 1988; **91**:177-183.
- Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Medicina Basada en la Evidencia. Cómo ejercer y enseñar la MBE. Ed. Churchill Livinstone, Madrid 1997.

- 7 Department of Clinical Epidemiology and Biostatistics. McMaster University Health Science Center. How to read clinical journals: II. To learn about a diagnostic test. *Can Med Assoc J* 1981; **124**:703-710.
- 8 Jaeschke R, Guyatt G, Sackett DL, for the Evidence-based Medicine Working Group. User's guide to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid?. *JAMA* 1994; **271**:389-391.
- 9 Jaeschke R, Guyatt G, Sackett DL, for the Evidence based Medicine Working Group. User's guide 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? *JAMA* 1994; **271**:703-707.
- 10 Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA* 1995; **1274**:645-651.
- 11 González de Dios J. La Medicina Basada en la Evidencia: ¿qué es?, ¿por qué?, ¿para qué?. *Pediatr Integral* 1.999; **3**: 493-497.
- 12 Burgueño MJ, García-Bastos JL, González-Buitrago JM. Las curvas ROC en la evaluación de pruebas diagnósticas. *Med Clin (Barc)* 1995; **104**:661-670.
- 13 Hernández I, García AM. La evaluación de pruebas diagnósticas en España. Una aproximación crítica. *Rev Salud Publ* 1993; **3**:243-262.
- 14 López Piñero JAM, Terrada ML. El consumo de información científica nacional y extranjera en las revistas médicas españolas: un nuevo repertorio destinado a su estudio. *Med Clin (Barc)* 1994; **102**:104-112.
- 15 González de Dios J, Moya M. Estudio bibliométrico de "Anales Españoles de Pediatría" (década 1984-1993) I: Análisis de los artículos publicados. *An Esp Pediatr* 1995; **42**: 2-10.
- 16 González de Dios J, Moya M. Estudio bibliométrico de "Anales Españoles de Pediatría" (década 1984-1993) II: Análisis de las referencias bibliográficas. *An Esp Pediatr* 1995; **42**: 11-18.
- 17 Robertson EA, Zweig MH, Van Steirghem AC. Evaluating the clinical efficacy of laboratory tests. *Am J Clin Pathol* 1982; **79**: 78-86.
- 18 Sheps SM, Schechter MT. The assessment of diagnostic tests. A survey of current medical research. *JAMA* 1984; **252**:2418-2422.
- 19 Arroll B, Schechter MT, Sheps SM. The assessment of diagnostic tests: a comparison of medical literature in 1982 and 1985. *J Gen Intern Med* 1988; **3**:443-447.
- 20 Mulrow CD, Linn WD, Gaul MK, Pugh JA. Assessing quality of a diagnostic test evaluation. *J Gen Intern Med* 1989; **4**:288-295.
- 21 Feinstein AR. Diagnostic and spectral markers. En: Clinical epidemiology: the architecture of clinical research. Filadelfia: W.B. Saunders 1985; 597-631.
- 22 González de Dios J, Moya M. Análisis bibliométrico de las publicaciones sobre Neonatología en "Anales Españoles de Pediatría" (período 1984-1996). *RELAN* 1998; **1**:12-25.
- 23 Begg CB. Bias in the assessment of diagnostic test. *Stat Med* 1987; **6**:411-423.
- 24 Bates AS, Margolis PA, Evans AT. Verification bias in pediatric studies evaluating diagnostic tests. *J Pediatr* 1993; **122**:585-590.
- 25 Greenes RA, Begg CB. Assessment of diagnostic technologies: methodology for unbiased estimation from samples of selectively verified patients. *Invest Radiol* 1985; **20**:751-756.
- 26 Philbrick JT, Horwitz RI, Feinstein AR. Methodologic problems of exercise testing for coronary artery disease: groups, analysis and bias. *Am J Cardiol* 1980; **46**:643-651.
- 27 Diamond GA, Rozanski A, Forrester JS, et al. A model for assessing the sensitivity and specificity of tests subject to selection bias. *I Chron Dis* 1986; **39**:206-215.
- 28 Lachs MS, Nachamkin Y, Edelstein PH, Goldman J, Feinstein AR, Schwartz JS. Spectrum bias in the evaluation of diagnostic tests: lessons from the rapid dipstick test for urinary tract infection. *Ann Intern Med* 1992; **117**:135-140.
- 29 De Mata Donado J, Orejas G. Epidemiología y metodología científica aplicada a la Pediatría (III) : Introducción al análisis estadístico en epidemiología. *An Esp Pediatr* 1999; **50**: 81-90.
- 30 Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br Med J* 1986; **292**: 746-750.
- 31 Porta Serra M, Plasencia A, Sanz F. La calidad de la información clínica (y III): ¿estadísticamente significativo o clínicamente importante? *Med Clin (Barc)* 1988; **90**:463-468.
- 32 González de Dios J. Niveles de calidad de la evidencia científica. *Pediatr Integral* 1994; **4**:108-113.
- 33 Ochoa Sangrador C, Orejas G. Epidemiología y metodología científica aplicada a la Pediatría (IV) : Pruebas diagnósticas. *An Esp Pediatr* 1999; **50**:301-314.

Artículos revisados

- 1 San Román L, Escobar M, García FJ, et al. Bilirrubinometría transcutánea en niños de bajo peso al nacimiento. *An Esp Pediatr* 1994; **40**:368-370.
- 2 Díez J, Calvo F, Martínez P, et al. Valor del hematocrito de cordón para el diagnóstico precoz de la policitemia neonatal. *An Esp Pediatr* 1994; **41**:89-92.
- 3 Armas H, González CM, González NL, Milena A, Doménech E. Screening neonatal de fibrosis quística mediante tripsina inmunorreactiva sérica. *An Esp Pediatr* 1994; **41**:261-266.
- 4 Villanustre C, Buznego R, Rodicio M, et al. Estudio comparativo de los métodos semicuantitativos (leucocituria, test de nitritos y Uricult) con el urocultivo para el diagnóstico de infección urinaria en el lactante. *An Esp Pediatr* 1994; **41**:325-328.
- 5 Vilanova JM, Corretger JM, Romera G, Figueras J, Vila J, Jiménez R. Utilidad clínica de la detección de antigenuria en el diagnóstico de la infección estreptocócica neonatal. *An Esp Pediatr* 1995; **42**:203-206.
- 6 González A, Sánchez M, Arnaiz P, et al. Historia familiar, práctica de deporte o tiempo dedicado a ver la televisión. ¿Cuál es el mejor predictor de la hipercolesterolemia en niños y adolescentes? *An Esp Pediatr* 1995; **42**:337-340.
- 7 Balaguer JA, Garbayo J, Bofarull JM, Fernández JD, Escribano J, Soler L. Hipoglucemias en recién nacidos a término macrosomáticos no proporcionados. Valor del índice ponderal para predecir el riesgo de hipoglucemia neonatal. *An Esp Pediatr* 1995; **43**:124-127.
- 8 Luaces C, García JJ, Garron P, Sanchez ML, Camacho E. Utilidad de un sistema de puntuación clínico y la pulsioximetría (SaO₂) en la valoración de la gravedad de las crisis de asma. *An Esp Pediatr* 1996; **44**:429-432.
- 9 Muñoz MA, Obregón E, Navarro J, Gurbindo MD, Hernández T, Fernández Cruz E. Estudio comparativo de técnicas para el diagnóstico del virus de la inmunodeficiencia humana en niños menores de 15 meses por: cultivo viral, reacción en cadena de la polimerasa y antígeno p24. *An Esp Pediatr* 1996; **44**:540-544.
- 10 Piñero E, Plaza J, Muñoz E. Valoración del análisis sistemático de orina en el lactante sano. *An Esp Pediatr* 1996; **45**:140-142.
- 11 Benito J, Sánchez J, Mintegui S, Montejo M. Infección urinaria en el lactante: Comprobación por punción suprapúbica del rendimiento del cultivo de orina obtenido por bolsa perineal. *An Esp Pediatr* 1996;

- 45:149-152.
- 12 Benito J, Mintegui S, Sánchez J, Montejó M, Soto E. Utilidad de la saturación de oxígeno y del pico máximo de flujo espiratorio en el manejo de las crisis asmáticas. *An Esp Pediatr* 1996; **45**:361-364.
 - 13 Gasser I, Vives D, Salcedo S, Olona M, Fernández F. Valoración de los cultivos superficiales del neonato con riesgo de infección precoz. Estudio de 735 pacientes. *An Esp Pediatr* 1996; **45**:409-414.
 - 14 Sequí JM, Tomás M, Marco J, Paredes C, Brines J. Diagnóstico de otitis secretoria mediante reflexometría acústica. *An Esp Pediatr* 1996; **45**:483-486.
 - 15 Panero A, Nogales A. Neopterina como marcador de la activación inmune en infecciones víricas y bacterianas. *An Esp Pediatr* 1996; **45**:597-602.
 - 16 Andreu A, Salcedo S, Heredia F, González J, Bartolomé RM, Cabero LL. Evaluación de tres técnicas rápidas para la detección intraparto del estreptococo del grupo B. *An Esp Pediatr* 1997; **46**:378-382.
 - 17 Sánchez M, Fernández ML. Colesterol HDL en la infancia. Su influencia en el diagnóstico de la hipercolesterolemia. *An Esp Pediatr* 1997; **47**:285-288.
 - 18 Muñoz C, Gené A, Pérez I, Mira A, Roca J, Latorre C. Diagnóstico de la tuberculosis en niños. Evaluación de la técnica reacción en cadena de la polimerasa. *An Esp Pediatr* 1997; **47**:353-356.
 - 19 Garrido M, Blanco A, Garrote JA, Tellería JJ, Arranz E. Valor de los anticuerpos salivales para la determinación de la seropositividad frente al sarampión, rubéola y parotiditis en niños y adultos. *An Esp Pediatr* 1997; **47**:499-504.
 - 20 Bonillo A, Romero J, Picazo B, Tapia L, Romero J, Martínez MI. Valor pronóstico y precisión de los indicadores de crisis asmática severa. *An Esp Pediatr* 1997; **47**:606-610.
 - 21 Galindo M, Calleja S, Nieto MA, Fadrique B, González AM, Manzanares J. Valor diagnóstico de la ecografía en la apendicitis del niño. *An Esp Pediatr* 1998; **48**:28-32.
 - 22 Calvo F, Sendra S, Mialaret A, Montagud E, Llanes S, Medrano J. Valor de la proteína C-reactiva en el diagnóstico de la apendicitis aguda en el niño. *An Esp Pediatr* 1998; **48**:376-380.
 - 23 Sánchez J, Luis M, Vázquez MA, Mintegui S, Benito J, López P. Valor diagnóstico de la proteína C reactiva en las sospechas de apendicitis aguda en la infancia. *An Esp Pediatr* 1998; **48**:470-474.
 - 24 Resino S, Alonso R, Jiménez JL, Gurbindo D, Muñoz MA. Determinación de la carga viral para el diagnóstico precoz de la infección perinatal por el virus de la inmunodeficiencia humana tipo 1 (VIH-1). *An Esp Pediatr* 1998; **49**:60-64.
 - 25 García de la Rubia S, Pajarón MJ, Martínez I, Sánchez M, Pérez D, Pajarón de Ahumada M. Comportamiento del FEVI y del PEFr en la prueba de la carrera libre para la detección del asma inducida por ejercicio en la infancia. *An Esp Pediatr* 1998; **49**: 237-240.
 - 26 Parejo R, Olivares F, Escobar H, Jiménez I, De Rafael L, Camarero C. Análisis comparativo de los métodos diagnósticos de la infección por *Helicobacter pylori* en el niño. *An Esp Pediatr* 1998; **49**:257-263.