

EVIDENCIAS EN PEDIATRÍA

Toma de decisiones clínicas basadas en las mejores pruebas científicas

www.evidenciasenpediatria.es

Fundamentos de medicina basada en la evidencia

Listas guía de comprobación de estudios sobre pruebas diagnósticas incluidos en las revisiones sistemáticas: declaración QUADAS

González Rodríguez MP¹, Velarde Mayol C²

¹CS Barrio del Pilar. Área 5. Madrid (España).

²CS de Segovia 1. Segovia (España).

Correspondencia: María Paz González Rodríguez, pazgonz@gmail.com

Palabras clave en inglés: diagnostic tests/routine standards; checklist: utilization; diagnosis; checklist: methods; checklist: standards.

Palabras clave en español: pruebas diagnósticas: estándares; listas guía: utilización; diagnóstico; listas guía: métodos; listas guía: estándares.

Fecha de recepción: 26 de febrero de 2012 • **Fecha de aceptación:** 27 de febrero de 2012

Fecha de publicación del artículo: 1 de marzo de 2012

Evid Pediatr.2012;8:20.

CÓMO CITAR ESTE ARTÍCULO

González Rodríguez MP, Velarde Mayol C. Listas guía de comprobación de estudios sobre pruebas diagnósticas incluidos en las revisiones sistemáticas: declaración QUADAS. Evid Pediatr. 2012;8:20.

Para recibir Evidencias en Pediatría en su correo electrónico debe darse de alta en nuestro boletín de novedades en <http://www.evidenciasenpediatria.es>

Este artículo está disponible en: <http://www.evidenciasenpediatria.es/EnlaceArticulo?ref=2012;8:20>

©2005-12 • ISSN: 1885-7388

Listas guía de comprobación de estudios sobre pruebas diagnósticas incluidos en las revisiones sistemáticas: declaración QUADAS

González Rodríguez MP¹, Velarde Mayol C²

¹CS Barrio del Pilar. Área 5. Madrid (España).

²CS de Segovia 1. Segovia (España).

Correspondencia: María Paz González Rodríguez, pazgonz@gmail.com

Cuando un clínico se enfrenta a la labor de interpretar una prueba diagnóstica, está tratando de dilucidar si un paciente tiene una determinada enfermedad o situación clínica, en qué estado evolutivo se encuentra y la gravedad que tiene¹. El conocimiento de los resultados de las pruebas diagnósticas, en combinación con la historia clínica y la exploración del paciente, le orientará en la toma de decisiones acerca del pronóstico y del tratamiento.

Por ello, determinar en qué medida la prueba diagnóstica utilizada mide lo que dice medir (validez) y si se puede utilizar en diferentes situaciones (fiabilidad) es de vital importancia.

El término prueba diagnóstica se refiere a cualquier procedimiento que puede proporcionar información acerca de la salud de una persona. Puede incluir análisis de laboratorio, exploraciones quirúrgicas, exámenes clínicos, pruebas de imagen, cuestionarios y estudios de anatomía patológica. Ante una nueva prueba diagnóstica y antes de utilizarla en la práctica clínica, es necesario conocer su validez.

Los estudios de validez de pruebas diagnósticas tienen el objetivo de determinar la exactitud o precisión de las pruebas diagnósticas que se evalúan y permitir el cálculo de variables estadísticas que proporcionan información de la eficacia y rendimiento de la prueba. Es decir, de lo apropiado que es la prueba en cuestión para detectar una enfermedad determinada. Así pues, para determinar la exactitud de una prueba diagnóstica, el resultado de la misma se compara con el resultado de la prueba de referencia, y se expresa en forma de sensibilidad, especificidad o combinaciones de estos índices, como los cocientes de probabilidad o los valores predictivos. En aquellas pruebas en que los resultados se pueden utilizar con distintos puntos de corte, el resultado se expresa como una curva ROC (curva de rendimiento diagnóstico), que refleja la sensibilidad y la especificidad para distintos puntos de corte.

Los criterios de validez en un estudio de un test diagnóstico se resumen en tres preguntas básicas, sobre las cuales se construirán los criterios de calidad:

1. ¿Se incluyó un espectro apropiado de pacientes similares a los que se aplicará la prueba en la práctica clínica? Es decir se incluye a los pacientes llamados de "incertidumbre diagnóstica".

2. ¿Hubo comparación ciega con una prueba de referencia (*gold standard*) independiente y adecuado? Es decir la prueba de estudio no puede formar parte de la prueba de referencia.

3. ¿Se evaluaron los resultados de la prueba de referencia independientemente de la prueba en estudio?

Al hablar de **calidad** de un estudio acerca de una prueba diagnóstica, podemos referirnos a varios aspectos. Para que los estudios de pruebas diagnósticas sean validos es preciso que estén basados en un rigor metodológico en cuanto al diseño y a la aplicabilidad.

En primer lugar interesa conocer si existen sesgos en la estimación de la validez de la prueba diagnóstica¹. Numerosos sesgos como el de verificación, incorporación o modificación del estado de la enfermedad, entre otros, son importantes a la hora de valorar la calidad de un estudio.

El segundo aspecto es acerca de cómo se realizó el estudio, en cuanto a tamaño de la muestra, análisis de subgrupos, si existía un protocolo preestablecido del estudio.

Un tercer aspecto es la calidad de la aplicabilidad de los resultados, también llamado validez externa, es decir, en qué medida se pueden generalizar a otros pacientes y en otros lugares. Dependerá de qué características tenían los pacientes incluidos en el estudio, el tipo de prueba utilizada y del ámbito en el que se desarrolló.

Otro aspecto es el de la descripción de los resultados, quiénes se incluyeron, cómo se hizo la prueba y la precisión de la misma, entre otros.

Existen numerosas herramientas de evaluación de la calidad de los estudios de pruebas diagnósticas. Entre las más utilizadas están las guías de usuario de la literatura médica³, basadas en las recomendaciones del Evidence-Based Medicine Working Group donde se describen los pasos a seguir para analizar la validez, la importancia y la aplicabilidad de los estudios sobre pruebas diagnósticas. Estos criterios han sido descritos en un artículo de *Evidencias Pediatría*⁴.

En una revisión sistemática acerca de las herramientas disponibles para evaluar los estudios de exactitud de pruebas diagnósticas, se encontraron hasta 91 escalas de valoración². Sin embar-

go, la mayoría de las escalas no incluían una definición de calidad y no habían sido evaluadas de forma sistemática.

Con el objetivo de establecer una herramienta para valorar la precisión de las pruebas diagnósticas se desarrolló la **declaración STARD**⁵ (Standard for Reporting of Diagnostic Accuracy). En 1999, el grupo de trabajo de la Cochrane de pruebas diagnósticas se reunió en Roma. El grupo diseñó una lista de 25 ítems mediante el consenso de expertos, con el objetivo de mejorar el diseño de los estudios que investigan la precisión diagnóstica de los test o pruebas. El grupo de trabajo siguió el ejemplo de la iniciativa CONSORT en el área de los ensayos clínicos, como se describió en la serie *Fundamentos de Evidencias en Pediatría*⁶. Es por así decirlo, en el área de la investigación sobre diagnóstico lo que CONSORT es en el área de ensayos clínicos.

Define un listado de 25 preguntas y un diagrama de flujo que debería seguirse para que el diseño de un estudio fuese adecuado, teniendo en cuenta la inclusión de los pacientes, el orden de la realización de la prueba, el número de pacientes que reciben la prueba y la prueba de referencia seleccionada. La iniciativa STARD está dirigida a editores de revistas y a los autores de artículos, con el objetivo de poder valorar los sesgos potenciales del estudio (validez interna) y la generalización o aplicabilidad de los resultados (validez externa). Se puede consultar en: <http://www.stard-statement.org/>

En las revisiones sistemáticas es importante la evaluación de la calidad de los estudios incluidos. Las revisiones sistemáticas de estudios de pruebas diagnósticas tienen una dificultad añadida a las de los estudios acerca de tratamientos. Los estudios son observacionales, susceptibles de introducir sesgos. Además, puede existir variabilidad en los métodos, proveedores, procedimientos y en las escalas que se utilizan para valorar la precisión de las pruebas⁷. En los estudios que analizan la precisión de las pruebas diagnósticas, se ha descrito la escasez de herramientas validadas para valorar la calidad de los mismos⁸. Con el objetivo de evaluar la calidad de los estudios de precisión de pruebas diagnósticas incluidos en las revisiones sistemáticas, se ha desarrollado la escala de valoración **QUADAS**⁹.

La declaración QUADAS paso a paso, ítem a ítem

La herramienta QUADAS (Quality Assessment Diagnostic Accuracy Studies) se desarrolló como un proyecto colaborativo entre el Centre for Reviews and Dissemination, University of York, y la Academic Medical Centre de la Universidad de Amsterdam. Fue financiado por el programa Health Technology Assessment (HTA) y se publicó en el año 2003.

Desde entonces ha sido utilizado en un gran número de revisiones sistemáticas. Una versión modificada del QUADAS ha sido utilizada por la Colaboración Cochrane en las revisiones de la precisión de pruebas diagnósticas (Diagnostic Test Accuracy Working Group)¹¹. Otros organismos que recomiendan su utilización son NICE y AHRQ. Un grupo de expertos diseñaron una lista de ítems relevantes, recogidos en la literatura médica. Utilizando el método Delphi seleccionaron 14 ítems. Cada uno se puntuaba como “sí”, “no” o “dudoso”. El “sí”

indicaba siempre una buena respuesta. El QUADAS incluye el riesgo de sesgo, aplicabilidad y calidad en la descripción del estudio. La versión Cochrane de la herramienta omitió los ítems relacionados con la calidad en la descripción del estudio.

A partir de la experiencia de los autores y las aportaciones de la Cochrane en cuanto a dificultades con la utilización de QUADAS, se procedió a revisar la primera versión y a desarrollar el QUADAS-2¹² en el año 2010.

El QUADAS-2¹³ está formado por cuatro áreas fundamentales que incluyen:

1. La selección de los pacientes.
2. El test o prueba en estudio.
3. Los estándares de referencia.
4. El flujo de los pacientes y el cronograma.

En cada una de las áreas se evalúa el riesgo de sesgo y las dudas acerca de su aplicabilidad. Esta evaluación se realiza con una serie de preguntas orientadas a evaluar existencia de un sesgo. El QUADAS-2 se aplica en cuatro fases:

1. La fase 1 es un resumen de la pregunta de la revisión. En esta fase, los autores deben describir las características de la revisión sistemática: los pacientes, la prueba diagnóstica, la prueba de referencia y la enfermedad o situación que se estudia.
2. La fase 2 incluye adaptar al estudio en concreto que se analiza, las preguntas orientativas que ayudarán a interpretar la calidad del estudio.
3. La fase 3 se refiere a establecer un diagrama de trabajo (diagrama de flujo). Se revisa el diagrama de flujo de cada estudio primario y si no lo tiene, se realiza un diagrama de cada estudio. Con un diagrama de trabajo apropiado será más fácil evaluar el riesgo de sesgo. De esta manera se obtiene información acerca de método de inclusión de los pacientes (es decir si son pacientes consecutivos con síntomas que hacen sospechar que tienen la enfermedad, o si son casos y controles).
4. En la fase 4 se valora el riesgo de sesgo y la aplicabilidad:
 - Riesgo de sesgo:

La primera parte de cada área se valora el sesgo y está estructurada en tres secciones: 1) qué información se proporciona para poder evaluar el riesgo de sesgo; 2) las preguntas orientativas, y 3) la valoración del riesgo de sesgo.

Las preguntas orientativas se responden con los términos: “sí”, “no”, o “dudoso”. El riesgo de sesgo se valora como “bajo”, “alto”, o “dudoso”. Si todas las preguntas orientativas son respondidas como sí, entonces el riesgo es bajo. Si alguna se responde como no, existe riesgo de sesgo. En ese caso los autores deben utilizar las guías desarrolladas en la fase dos para juzgar el riesgo de sesgo.
 - Aplicabilidad: los autores de la revisión deben registrar la información en base a lo que se concluye acerca de la aplicabilidad.

Tabla. EVALUACIÓN DE LA CALIDAD (RIGOR METODOLÓGICO) DE LOS ESTUDIOS DE PRUEBAS DIAGNÓSTICAS.

Área	Selección de los pacientes	Prueba diagnóstica en estudio	Prueba de referencia	Flujo y cronograma
Descripción	Describe los métodos utilizados para seleccionar a los pacientes: pruebas previas, ámbito, uso previsto de la prueba en estudio	Describe la prueba, cómo se realizó y su interpretación	Describe la prueba de referencia, cómo se realizó y su interpretación	Describe a los pacientes que no van a recibir la prueba de estudio, la prueba de referencia o que se excluyen de la tabla 2 x 2: describe el intervalo y cualquier intervención entre la prueba en estudio y la de referencia
Preguntas clave (sí/no/dudoso)	<p>¿Es una muestra consecutiva o aleatoria?</p> <p>¿Se evitó un diseño de casos y controles?</p> <p>¿Se evitaron exclusiones inapropiadas?</p>	<p>¿Se interpretaron los resultados de la prueba sin el conocimiento de los de la prueba de referencia? Lo correcto es realizar primero la prueba de estudio</p> <p>Si se usó un punto de corte (umbral), ¿se especificó previamente?</p>	<p>¿La prueba de referencia clasifica correctamente la enfermedad en estudio?</p> <p>¿Los resultados de la prueba de referencia se interpretaron independientemente de la prueba de estudio?</p> <p>¿Hay algún elemento de la prueba en estudio que forme parte de la prueba de referencia?</p>	<p>¿Describe el intervalo de tiempo entre las dos pruebas?</p> <p>¿El intervalo de tiempo es el adecuado?</p> <p>¿Se aplicó a todos los pacientes el patrón de referencia?</p> <p>¿Todos los pacientes recibieron la misma prueba de referencia independientemente del resultado de la prueba en estudio?</p> <p>¿Se incluyeron todos los pacientes en el análisis?</p>
Riesgo de sesgo (alto/bajo/dudoso)	¿Hay sesgo en la selección de los pacientes?	¿Podría haber sesgos en la realización e interpretación de la prueba?	¿Podría haber sesgos en la realización e interpretación de la prueba?	¿El flujo de seguimiento del paciente podría haber producido algún sesgo?
Aplicabilidad (alta/baja/dudosa)	¿Hay dudas de que los pacientes incluidos y su ámbito de estudio no se ajusten a la pregunta de la revisión? Es decir, que sean diferentes de la población diana	¿Hay dudas de que la prueba (realización e interpretación) difieran de la pregunta de revisión? Cualquier modificación de la tecnología, interpretación o realización merma su aplicabilidad	¿Hay dudas de que a condición de estudio (enfermedad) definida por la prueba de referencia (realización e interpretación) difiera o no se ajustara a la pregunta de revisión?	

Fuente: Table that summarises QUADAS-2 and lists all signalling, risk of bias and applicability rating questions. (fecha de consulta 26 febrero 2012). Disponible en: <http://www.bris.ac.uk/quadas/quadas-2>

La versión de QUADAS-2 se puede descargar de la página web de la University of Bristol: <http://www.bris.ac.uk/quadas/quadas-2/>

La valoración del riesgo de sesgo y de la aplicabilidad se puede consultar en formato de tabla en: <http://www.bris.ac.uk/quadas/quadas-2/>

ÁREAS FUNDAMENTALES

La selección de los pacientes

- Riesgo de sesgo: ¿podría haber habido sesgo en la selección de los pacientes?

Para responder a esta pregunta se incluyen las siguientes preguntas orientativas:

1. ¿Se incluyó una muestra de pacientes consecutiva o aleatoria?
2. ¿Se evitó un diseño de casos y controles?
3. ¿Se evitó en el estudio que hubiese exclusiones inapropiadas?

Un estudio ideal debe incluir todos los pacientes consecutivos, o una muestra aleatoria de los pacientes con la sospecha de la enfermedad. Si un estudio excluye pacientes de forma inapropiada, proporcionará unos resultados que sobrestimarán la precisión.

- Aplicabilidad: ¿hay dudas de que los pacientes incluidos y el lugar no se ajusten a la pregunta de la revisión? Los pacientes pueden ser diferentes en cuanto a la condición inicial, las características demográficas etc. Si los pacien-

tes del estudio son diferentes de la población definida en la pregunta de la revisión, puede haber dudas acerca de la aplicabilidad.

Prueba de estudio

- **Riesgo de sesgo:** ¿se puede haber producido algún sesgo al interpretar la prueba en estudio?
Para responder a esto se incluyen las siguientes preguntas orientativas:
 1. ¿Se realizó la interpretación de la prueba en estudio sin conocer los resultados de la prueba de referencia?
Este ítem es parecido al cegamiento de los estudios de intervención. El conocimiento de los resultados de la prueba de referencias puede influenciar la interpretación de los resultados de la prueba en estudio. También tiene importancia el orden de la realización de las pruebas. Si la prueba en estudio se hace y se interpreta antes de la de referencia entonces la persona no sabe los resultados de la prueba de referencia y se puede contestar como “sí”.
 2. Si se utilizó un punto de corte, ¿éste se especificó previamente?
Si se determina un punto de corte para obtener una sensibilidad y especificidad máxima, la prueba puede producir unos resultados muy optimistas de la eficacia de la prueba.
- **Aplicabilidad:** ¿existen dudas que la prueba de estudio, su realización, o interpretación difieran de la pregunta de la revisión?
Variaciones en las tecnologías, realización de la prueba y la interpretación afecta a la precisión y de esta manera a la aplicabilidad.

Prueba de referencia

- **Riesgo de sesgo:** ¿puede existir algún sesgo en el estándar de referencia, su realización o en su interpretación?
 1. ¿El estándar de referencia clasifica correctamente la enfermedad estudiada?
El cálculo del rendimiento de una prueba se basa en asumir que el estándar de referencia tiene una sensibilidad y una especificidad del 100%. Cualquier desacuerdo se asume que procede de una clasificación incorrecta de la prueba en estudio.
Por ello es muy importante que los autores de la revisión tengan seguridad acerca de que si el estándar de referencia es positivo, pues es que tienen la enfermedad, y si es negativo, no la tienen.
 2. ¿Se procedió a interpretar los resultados del estándar de referencia sin conocer los resultados de la prueba en estudio?
Esto es similar al cegamiento de los estudios de intervención.
- **Aplicabilidad:** ¿existen dudas acerca de que la enfermedad en estudio definida por la prueba de referencia difiera de la pregunta de la revisión?

Flujo y cronograma

- **Riesgo de sesgo:** ¿puede que el flujo del paciente haya introducido un sesgo?
 1. ¿El intervalo entre la prueba de estudio y la prueba de referencia fue adecuado?
En una situación ideal, ambas pruebas se deberían realizar en el mismo paciente al mismo tiempo. Si se produce un retraso o si ha empezado el tratamiento entre ambas, puede producirse un error en la clasificación. Una separación de unos días en una enfermedad crónica puede no ser importante. En cambio en una enfermedad infecciosa puede ser un problema.
 2. ¿A todos los pacientes se les realizó la misma prueba de referencia?
En el caso de que los resultados de la prueba en estudio condicione la realización la prueba de referencia, se produce el llamado sesgo de verificación.
 3. ¿Se incluyeron todos los pacientes en el análisis?
El análisis final se debe realizar con todos los pacientes incluidos al comienzo del estudio, ya que las pérdidas durante el seguimiento pueden alterar la interpretación de los resultados.

Como consideraciones finales, QUADAS-2 no debe utilizarse para generar una escala de puntuación de la calidad. Si un estudio se considera como “bajo” en todas las áreas pues el estudio se describe como “bajo riesgo de sesgo”. Si “alto” o “dudoso” pues se describe como “riesgo de sesgo” o “dudas acerca de la aplicabilidad”.

Los resultados se pueden describir en forma de resumen o de tabla, describiendo cuantos estudios tienen bajo, alto o dudoso riesgo de sesgo y aplicabilidad en cada área.

Los autores pueden elegir incluir únicamente los estudios de pruebas diagnósticas con bajo riesgo de sesgo en todas las áreas. También pueden realizar análisis de subgrupos y análisis de sensibilidad.

Así pues, a la hora de realizar una revisión sistemática de estudios de pruebas diagnósticas, es preciso realizar una valoración detallada de la calidad de los estudios incluidos. La herramienta QUADAS-2 se utiliza para valorar la calidad de los estudios.

Aunque no es el objetivo de este artículo el realizar una comparación entre QUADAS y STARD, a continuación se describen algunos aspectos que pueden ser de utilidad para el lector. Ambas iniciativas coinciden en la búsqueda de un instrumento que detecte la variación y el sesgo de los estudios de pruebas diagnósticas utilizando la Medicina basada en la evidencia.

Difieren entre sí en la intención del instrumento: STARD tiene como objetivo el proporcionar una lista que sirva de guía para la publicación de los estudios de precisión de pruebas diagnósticas. Es una herramienta que se utiliza de forma prospectiva para realizar un diseño adecuado de un estudio; por tanto, interesan a los investigadores en la fase de diseño del estudio y a los editores. QUADAS-2 es una herramienta para valorar la

calidad de los estudios primarios en las revisiones sistemáticas y metaanálisis. Se utiliza de forma retrospectiva para realizar un análisis crítico del rigor metodológico de un estudio de pruebas diagnósticas.

Los puntos críticos del QUADAS recaen en su reproductibilidad, sobre todo los ítems relativos a los resultados indeterminados o no concluyentes, las pérdidas y retiradas del estudio.

No obstante, algunos autores recomiendan utilizar ambas escalas a la hora de evaluar la calidad de los estudios de pruebas diagnósticas¹⁵. Este tema será tratado en otro artículo de la sección de *Fundamentos de Evidencias en Pediatría*.

BIBLIOGRAFÍA

1. Zamora J, Abaira V. Análisis de la calidad de los estudios de evaluación de pruebas diagnósticas. *Nefrología*. 2008; 28:42-5.
2. Whiting P, Rutjes AWS, Dinnes J, Reitsma JB, Bossuyt P, Kleijnen J. A systematic finds that diagnostic reviews fail to incorporate quality despite available tools. *J Clin Epidemiol*. 2005;58:1-12.
3. 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.
4. Ochoa Sangrador C, González de Dios J, Buñuel Álvarez JC. Evaluación de artículos científicos sobre pruebas diagnósticas. *Evid Pediatr*. 2007;3:24.
5. STARD Statement. Standards for the Reporting of Diagnostic accuracy Studies. Fecha de consulta 26 de febrero de 2012. Disponible en: <http://www.stard-statement.org/>
6. González de Dios J, Buñuel Álvarez JC, González Rodríguez P. Listas guía de comprobación de ensayos clínicos: declaración CONSORT. *Evid Pediatr*. 2011;7:72.
7. Mallett S, Deeks JJ, Halligan S, Hopewell S, Cornelius V, Altman DG. Systematic reviews of diagnostic tests in cancer: review of methods and reporting. *BMJ*. 2006;33:413.
8. Mallen CH, Peat G, Croft P. Quality assessment of observational studies is not common place in systematic reviews. *J Clin Epidemiol*. 2006;59:765-9.
9. Whiting P, Rutjes AW, Dinnes J, Reitsma J, Bossuit PM, Kleijnen J. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Assess*. 2004;8:No. 25.
10. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PM, Kleijnen J. The development Of. QUADAS: a tool for the quality assessment of Studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology*. 2003;3:25.
11. Cochrane Colaboration. Diagnostic Test Accuracy Working Group. Disponible en: <http://srdta.cochrane.org/welcome>
12. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy Studies. *Ann Intern Med*. 2011;155:529-36.
13. QUADAS. A quality assessment tool for diagnostic accuracy studies. University of Bristol. Fecha de consulta 19 febrero 2012. Disponible en: <http://www.bris.ac.uk/quadas/>
14. Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. *BMC Medical Research Methodology*. 2005;5:19.
15. Fernández de Oliveira MR, De Castro Gomes A, Toscano CM. QUADAS and STARD: evaluating the quality of diagnostic accuracy studies. *Rev Saúde Pública*. 2011;45.