

Comunicación científica (XXI). Conocimientos básicos para leer (y escribir) un artículo científico (8): Listas de comprobación de pruebas diagnósticas

J. González de Dios¹, M. González-Muñoz², A. Alonso-Arroyo³, R. Aleixandre-Benavent⁴

¹Servicio de Pediatría. Hospital General Universitario de Alicante. Departamento de Pediatría. Universidad «Miguel Hernández». Alicante. ²Facultad de Medicina y Odontología. Universidad de Valencia. ³Departamento de Historia de la Ciencia y Documentación. Facultad de Medicina y Odontología. Universidad de Valencia. ⁴Instituto de Gestión de la Innovación y del Conocimiento (CSIC-Universitat Politècnica de València). UISYS-Universitat de València

Resumen

Ante una nueva prueba diagnóstica, y antes de utilizarla en la práctica clínica, es necesario conocer su validez. Y para que los estudios de pruebas diagnósticas sean válidos, es preciso que tengan rigor metodológico en cuanto a su diseño y aplicabilidad. En las pruebas diagnósticas podemos considerar dos listas guía de comprobación: STARD para la validez y QUADAS para la calidad.

La declaración STARD define un listado de 25 preguntas y un diagrama de flujo que debería seguirse para que el diseño de un estudio sea 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 declaración QUADAS está formada por cuatro áreas fundamentales: a) selección de los pacientes; b) prueba en estudio; c) estándares de referencia, y d) flujo de pacientes y cronograma.

©2015 Ediciones Mayo, S.A. Todos los derechos reservados.

Palabras clave

Formación, información, listas de comprobación, medicina, pediatría, pruebas diagnósticas, publicación científica

Pruebas diagnósticas: entre la calidad y la precisión

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.

Cuando un clínico se enfrenta a la labor de interpretar una prueba diagnóstica, está tratando de dilucidar si un paciente tie-

Abstract

Title: Scientific communication (XXI). Basic knowledge to read (and write) a scientific paper (8): Checklists of diagnostic test

Before a new diagnostic test and before the use in clinical practice, it is necessary to know its validity. And to validate studies of diagnostic tests it is necessary they have methodological rigor in the design and applicability. In diagnostic tests we can consider two checklists: STARD for validity and QUADAS for quality.

The STARD statement defines a list of 25 questions and one flowchart to be followed for an appropriate study design, given the inclusion of patients, the order of performing the test, the number of patients receiving the test and the selected reference.

The QUADAS statement consists of four key areas including: selection of patients, test under study, reference standards, and patient flow and timetable.

©2015 Ediciones Mayo, S.A. All rights reserved.

Keywords

Checklist, diagnostic test, information, medicine, Pediatrics, scientific publication

ne una determinada enfermedad o situación clínica, en qué estado evolutivo está y la gravedad que reviste. El conocimiento de los resultados de las pruebas diagnósticas, en combinación con la historia clínica y la exploración del paciente, orientará al médico en la toma de decisiones acerca del pronóstico y el tratamiento. Por ello, es de vital importancia determinar en qué medida la prueba diagnóstica utilizada mide lo que dice medir (validez) y si se puede utilizar en diferentes situaciones (fiabilidad)¹.

Ante una nueva prueba diagnóstica, y antes de utilizarla en la práctica clínica, es necesario conocer su validez. Los estu-

dios de validez de pruebas diagnósticas tienen el objetivo de determinar la exactitud o precisión de las pruebas diagnósticas que se evalúan, así como permitir el cálculo de variables estadísticas que proporcionan información sobre la eficacia y el rendimiento de la prueba, es decir, lo apropiada que es la prueba en cuestión para detectar una enfermedad determinada. Así pues, para determinar la exactitud de una prueba diagnóstica, su resultado se compara con el de la prueba de referencia, y se expresa en forma de sensibilidad, especificidad o combinaciones de ambos índices, como los cocientes de probabilidad o los valores predictivos. En las pruebas en que los resultados se pueden utilizar con distintos puntos de corte, el resultado se expresa como una curva ROC (acrónimo de *receiver operating characteristic*), que refleja la sensibilidad y la especificidad para distintos puntos de corte².

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 válidos, es preciso que ofrezcan rigor metodológico en cuanto a su diseño y aplicabilidad:

- En primer lugar, interesa conocer si existen sesgos en la estimación de la validez de la prueba diagnóstica. 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 atañe a cómo se realizó el estudio en cuanto a tamaño de la muestra, análisis de subgrupos y si existía un protocolo preestablecido del estudio.
- Un tercer aspecto está relacionado con la calidad de la aplicabilidad de los resultados, también llamado validez externa; es decir, en qué medida éstos se pueden generalizar a otros pacientes y en otros lugares. Dependerá de qué características tengan los pacientes incluidos en el estudio, el tipo de prueba utilizada y el ámbito en que se desarrolló.
- Otro aspecto es el de la descripción de los resultados: qué pacientes se incluyeron en el estudio, cómo se realizó la prueba y su precisión...

En las pruebas diagnósticas podemos considerar dos listas guía de comprobación: STARD para la validez y QUADAS para la calidad.

Declaración STARD: prototipo de lista guía de comprobación de la validez de pruebas diagnósticas

Con el objetivo de establecer una herramienta para valorar la precisión de las pruebas diagnósticas se desarrolló la declaración STARD (Standard for Reporting of Diagnostic Accuracy)^{3,4}. En 1999, el grupo de trabajo de la Cochrane de pruebas diagnósticas se reunió en Roma y 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 clínicos⁵. Por

así decirlo, esta declaración representa en el área de la investigación sobre diagnóstico lo que CONSORT 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 sea 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 los 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).

Los resultados se publicaron en 2003 en importantes revistas médicas, como *Annals of Internal Medicine*, *Radiology*, *Clinical Chemistry* y *The British Medical Journal*⁶. Varias editoriales recomendaban a los autores y editores su aceptación. Su aprobación se basó en que la evaluación de la investigación depende de un informe completo y preciso, y que la incorporación de STARD contribuiría a mejorar el proceso^{6,7}. Su utilización fue impulsada por el Comité Internacional de Editores de Revistas Médicas, que en 2006 estableció su uso como requisito para la publicación de los trabajos.

Sin embargo, su aceptación ha sido variable⁸ y, actualmente, los criterios de calidad de STARD se cumplen en un porcentaje bajo de publicaciones^{9,10}. Esto no quiere decir que los estudios sean de mala calidad, sino que es necesario describir el proceso para que se puedan interpretar con transparencia los resultados.

Declaración STARD paso a paso, ítem a ítem

La declaración STARD clasifica la evaluación en cinco secciones, que son las clásicas IMRD de cualquier artículo original (Introducción, Métodos, Resultados y Discusión), con una sección previa (Título y Resumen). Está disponible una descripción detallada de cada punto en la página web <http://www.stard-statement.org/>¹¹.

Título y Resumen

1. Título. Identifica el artículo como un estudio de precisión diagnóstica (se recomienda emplear el término MeSH «sensibilidad y especificidad»).

Introducción

2. Especifica que las preguntas del estudio, o su objetivo, es estimar la precisión diagnóstica o comparar la precisión entre diferentes grupos y mediante otros test.

Métodos

Participantes

3. Describe la población incluida en el estudio: los criterios de inclusión y exclusión, y el lugar y la localización en que se realizó la recogida de datos.

- Describe cómo se seleccionaron los participantes: ¿la selección se basó en los síntomas iniciales, en los resultados de las pruebas previas, o en el hecho de que los participantes recibiesen la prueba de estudio y la prueba de referencia?
- Describe la selección de los participantes (el muestreo): ¿la población del estudio fue una serie consecutiva de participantes definida por los criterios de selección de los puntos 3 y 4? Si no es así, debe especificarse cómo se seleccionaron los participantes.
- Describe la recogida de datos: si se planeó previamente antes de realizar la prueba de estudio y la de referencia (estudio prospectivo) o después (estudio retrospectivo).

Métodos de las pruebas

- Describe la prueba de referencia y en qué se basa (su justificación).
- Describe las cuestiones técnicas del material y métodos, incluyendo cómo y cuándo se tomaron las medidas, y cita referencias para la prueba en estudio y la prueba de referencia.
- Describe la definición y las bases de las unidades, los puntos de corte y las categorías de los resultados de la prueba en estudio y de la prueba de referencia.
- Describe el número, la preparación y el grado de conocimiento de las personas que realizan e interpretan la prueba de estudio y la de referencia.
- Describe si la aplicación de la prueba y del test de referencia se realizó de forma ciega.

Métodos estadísticos

- Describe los métodos para calcular o comparar las medidas de precisión diagnóstica y las pruebas estadísticas utilizadas para cuantificar la incertidumbre (p. ej., los intervalos de confianza [IC] del 95%).
- Describe los métodos para calcular la reproducibilidad de la prueba, si éstos se realizaron (es decir, su fiabilidad).

Resultados

Participantes

- Describe cuándo se realizó un estudio, estableciendo las fechas de inicio y finalización de inclusión de los participantes.
- Describe las características clínicas y demográficas de la población de estudio (p. ej., edad, sexo, espectro de los síntomas, comorbilidad, tratamientos actuales y centros de reclutamiento).
- Describe el número de participantes que cumplen los criterios de inclusión y que recibieron o no las pruebas de estudio y la prueba de referencia; describe por qué los participantes fallaron en recibir alguna de las pruebas (se recomienda incluir un diagrama de flujo que identifique las pérdidas).

Resultados de las pruebas

- Describe el intervalo de tiempo entre la prueba de estudio y la prueba de referencia, y si se ha administrado algún tratamiento entre medias.

- Describe la distribución de la gravedad de la enfermedad (criterios definidos) en los pacientes con la enfermedad objeto de estudio, así como otros diagnósticos en los participantes sin la enfermedad objeto de estudio.
- Describe los resultados en una tabla (incluidos los indeterminados y los no obtenidos) a partir de los resultados de la prueba de estudio y la de referencia. En las variables continuas, la distribución de los resultados se establece según los valores de la prueba de referencia.
- Describe cualquier efecto adverso que pueda haber originado la realización de la prueba de estudio o la de referencia.

Estimaciones

- Describe las estimaciones de precisión diagnóstica y las medidas de incertidumbre estadística (p. ej., el IC del 95%).
- Describe cómo se interpretaron los resultados indeterminados, las pérdidas y los valores atípicos o extremos de la prueba de estudio.
- Describe la estimación de la variabilidad de la precisión diagnóstica, si ésta se llevó a cabo, entre subgrupos de participantes, investigadores o centros.
- Si procede, describe la estimación de la reproducibilidad (fiabilidad) de la prueba.

Comentario (Discusión)

- Comenta la aplicabilidad clínica de los resultados del estudio.

Declaración QUADAS: prototipo de lista guía de comprobación de la calidad de pruebas diagnósticas

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

Desde entonces, se ha utilizado en un gran número de revisiones sistemáticas. La Colaboración Cochrane ha utilizado una versión modificada del QUADAS en las revisiones de la precisión de pruebas diagnósticas (Diagnostic Test Accuracy Working Group)¹². Otros organismos que recomiendan su utilización son el National Institute for Health and Clinical Excellence (NICE) y la Agency for Healthcare Research and Quality (AHRQ).

El QUADAS original incluía 14 ítems y 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 esta 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 respecto a las dificultades en la utilización del QUADAS, se procedió a revisar la primera versión y a desarrollar el QUADAS-2¹³ en el año 2010.

El QUADAS-2¹⁴ está formado por cuatro áreas fundamentales: a) selección de los pacientes; b) prueba en estudio; c) estándares de referencia, y d) flujo de pacientes y 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 la existencia de un sesgo.

El QUADAS-2 se aplica en cuatro fases:

- 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: pacientes, prueba diagnóstica, prueba de referencia y enfermedad o situación que se estudia.
- La fase 2 implica adaptar al estudio en concreto que se analiza las preguntas orientativas que ayudarán a interpretar la calidad del estudio.
- En la fase 3 se establece 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 del 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).
- En la fase 4 se valora el riesgo de sesgo y la aplicabilidad:
 - Riesgo de sesgo. En 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) preguntas orientativas, y 3) 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», 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 2 para juzgar el riesgo de sesgo.
 - Aplicabilidad. Los autores de la revisión deben registrar la información a partir de la cual se concluye acerca de la aplicabilidad.

Declaración QUADAS paso a paso, ítem a ítem

A continuación se exponen las características más interesantes de cada ítem (toda la información se encuentra disponible en su web: <http://www.bris.ac.uk/quadas/quadas-2/>)¹⁵:

Selección de los pacientes

- Descripción. Describe los métodos utilizados para seleccionar a los pacientes: pruebas previas, ámbito, uso previsto de la prueba de estudio...

- Preguntas clave (sí/no/dudoso):
 - ¿Se incluyó una muestra de pacientes consecutiva o aleatoria?
 - ¿Se evitó un diseño de casos y controles?
 - ¿Se evitó en el estudio que hubiese exclusiones inapropiadas?
- Riesgo de sesgo (alto/bajo/dudoso):
 - ¿Hay sesgo en la selección de los pacientes?
- Aplicabilidad:
 - ¿Hay dudas de que los pacientes incluidos y el lugar no se ajusten a la pregunta de la revisión, es decir, que sean diferentes a la población diana?

Prueba diagnóstica en estudio

- Descripción. Describe la prueba, cómo se realizó y su interpretación.
- Preguntas clave (sí/no/dudoso):
 - ¿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. Este ítem es parecido al enmascaramiento de los estudios de intervención. El conocimiento de los resultados de la prueba de referencia puede influir en 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, no se conocen los resultados de esta última y se puede contestar como «sí».
 - Si se utilizó un punto de corte (umbral), ¿éste se especificó previamente? Si se determina un punto de corte para obtener una sensibilidad y una especificidad máximas, la prueba puede producir unos resultados muy optimistas acerca de la eficacia de la prueba.
- Riesgo de sesgo (alto/bajo/dudoso):
 - ¿Podría haber sesgo en la realización e interpretación de la prueba?
- Aplicabilidad:
 - ¿Hay dudas de que la prueba de estudio, su realización o interpretación difieran de la pregunta de la revisión? Las variaciones en las tecnologías, la realización de la prueba y la interpretación afectan a la precisión y, por tanto, a la aplicabilidad.

Prueba de referencia

- Descripción. Describe la prueba de referencia, cómo se realizó y su interpretación.
- Preguntas clave (sí/no/dudoso):
 - ¿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 resulta positivo es porque los pacientes tienen la enfermedad y si es negativo es porque no la tienen.

- ¿Se procedió a interpretar los resultados del estándar de referencia sin conocer los resultados de la prueba en estudio? ¿Hay algún elemento de la prueba de estudio que forme parte de la prueba de referencia?
- Riesgo de sesgo (alto/bajo/dudoso):
- ¿Podría haber sesgo en la realización e interpretación de la prueba?
- Aplicabilidad:
- ¿Hay dudas acerca de que la enfermedad en estudio definida por la prueba de referencia difiera de la pregunta de la revisión?

Flujo y cronograma

- Descripción. 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 × 2: describe el intervalo y cualquier intervención entre la prueba de estudio y la de referencia.
- Preguntas clave (sí/no/dudoso):
- ¿Describe el intervalo de tiempo entre las dos pruebas? ¿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.
- ¿A todos los pacientes se les realizó la misma prueba de referencia? En el caso de que los resultados de la prueba en estudio condicionen la realización de la prueba de referencia, se produce el llamado «sesgo de verificación».
- ¿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.
- Riesgo de sesgo (alto/bajo/dudoso):
- ¿El flujo de seguimiento del paciente podría haber producido algún sesgo?

Como consideraciones finales, el QUADAS-2 no debe utilizarse para generar una escala de puntuación¹ de la calidad. Si un estudio se considera como «bajo» en todas las áreas, se clasifica como de «bajo riesgo de sesgo»; si se considera «alto» o «dudoso», se clasifica como de «riesgo de sesgo» o con «dudas acerca de su aplicabilidad».

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 llevar a cabo una valoración detallada de la calidad de los estudios incluidos. La herramienta QUADAS-2 se utiliza para valorar la calidad de los estudios.

TABLA 1

Diferencias entre las declaraciones STARD y QUADAS-2

| STARD | QUADAS-2 |
|--|--|
| Para valorar la precisión de un estudio de pruebas diagnósticas | Para evaluar la calidad de una prueba diagnóstica |
| Aspecto que evaluar: la precisión, es decir, la capacidad de una prueba para identificar a los pacientes con la enfermedad | Aspecto que evaluar: el rigor de los estudios primarios |
| Su uso es prospectivo: <ul style="list-style-type: none"> • Para los investigadores en la fase de diseño del estudio • Para los editores | Su uso es retrospectivo: <ul style="list-style-type: none"> • Para las revisiones sistemáticas • Para los metaanálisis |
| Evalúa principalmente la validez interna: sesgos y variaciones | Evalúa principalmente su aplicabilidad |

Diferencias entre STARD y QUADAS

En resumen, la escala STARD se utiliza para valorar la precisión diagnóstica, es decir, para discernir la capacidad de una prueba de diferenciar entre pacientes que tienen la enfermedad de los que no la tienen. La herramienta STARD no está diseñada para valorar la calidad metodológica de los estudios de precisión diagnóstica; para ello, se debe utilizar otra lista de calidad, como la declaración QUADAS. 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 (tabla 1):

- El STARD tiene como objetivo 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, interesa a los investigadores en la fase de diseño del estudio y a los editores¹¹.
- El QUADAS es una herramienta para valorar la calidad de los estudios primarios en las revisiones sistemáticas y los metaanálisis. Se utiliza de forma retrospectiva para realizar un análisis crítico del rigor metodológico de un estudio de pruebas diagnósticas. Algunos autores recomiendan que la combinación de ambas aportaría mayor rigor a la hora de validar los estudios de pruebas diagnósticas¹².

Bibliografía

1. Zamora J, Abraira V. Análisis de la calidad de los estudios de evaluación de pruebas diagnósticas. *Nefrología*. 2008; 28: 42-45.
2. Ochoa Sangrador C, Buñuel Álvarez JC, González de Dios J. Evaluación de artículos sobre intervenciones diagnósticas. *Evid Pediatr*. 2007; 3: 24.
3. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwing LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ*. 2003; 326: 41-44.

4. Rennie D. Improving reports of studies of diagnostic tests. The STARD initiative. *JAMA*. 2003; 289: 89-90.
5. González de Dios J, González Muñoz M, Alonso Arroyo A, Aleixandre Benavent R. Comunicación científica (XIX). Conocimientos básicos para leer (y escribir) un artículo científico (6): Lista de comprobación de artículos experimentales. *Acta Pediatr Esp*. 2015; 73: 19-24.
6. Altman DG, Bossuyt PMM. Estudios de precisión diagnóstica (STARD) y pronóstica (REMARK). *Med Clin (Barc)*. 2005; 125: 49-55.
7. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med*. 2003; 138: 1-12.
8. Smidt N, Rutjes AW, Van der Windt DAWM, Ostelo RW, Bossuyt PM, Reitsma JB, et al. The quality of diagnostic accuracy studies since the STARD statement. Has it improved? *Neurology*. 2006; 67: 792-797.
9. Selman TJ, Morris RK, Zamora J, Khan KS. The quality of reporting of primary test accuracy studies in obstetrics and gynaecology: application of the STARD criteria. *BMC Womens Health*. 2011; 11: 8.
10. Gómez Sáez NG, Hernández-Aguado I, Lumberras B. Estudio observacional: evaluación de la calidad metodológica de la investigación diagnóstica en España tras la publicación de la guía STARD. *Med Clin (Barc)*. 2009; 133: 302-310.
11. González Rodríguez MP, Velarde Mayol C. Lista de comprobación de estudios sobre precisión de pruebas diagnósticas: declaración STARD. *Evid Pediatr*. 2012; 8: 43.
12. Cochrane Collaboration. Diagnostic Test Accuracy Working Group [consultado el 11 de marzo de 2015]. Disponible en: <http://srdta.cochrane.org/welcome>
13. 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: 529-536.
14. QUADAS. A quality assessment tool for diagnostic accuracy studies. University of Bristol [consultado el 11 de marzo de 2015]. Disponible en: <http://www.bris.ac.uk/quadas/>
15. 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.