

Valoración de la metodología de la investigación en pruebas de genética, molecular y proteómica

Blanca Lumbreras / Inmaculada Jarrín / Ildefonso Hernández Aguado

Departamento de Salud Pública, Historia de la Ciencia y Ginecología, Universidad Miguel Hernández, San Juan de Alicante, España.

(Evaluation of the research methodology in genetic, molecular and proteomic tests)

Resumen

Introducción: El avance en las tecnologías del análisis genómico ha supuesto el desarrollo de nuevas pruebas diagnósticas con potencial aplicación clínica. Así como ha sucedido en otros campos del diagnóstico, conocer las limitaciones metodológicas de la investigación en genética facilitará la aplicación de sus resultados.

Métodos: Se analizaron 44 artículos publicados en *JAMA*, *Lancet*, *New England Journal of Medicine*, *Cancer Research* y *Clinical Cancer Research*, de 2002 a junio de 2005, que estudiaban la exactitud diagnóstica de pruebas genéticas, moleculares y de proteómica. Se aplicaron 24 criterios metodológicos de la guía STARD (Standards for Reporting of Diagnostic Accuracy).

Resultados: La media de cumplimiento de los criterios aplicados fue de 9,8 (intervalo de confianza [IC] del 95%, 8,8-10,6), y las mayores deficiencias se observaron en los aspectos relacionados con la descripción de la selección de los pacientes (9; 20%), el tratamiento de los resultados indeterminados (5; 11%) y la determinación de la reproducibilidad de la técnica (6; 13%). En cambio, sí se observó una alta frecuencia en la descripción del patrón de referencia (39; 87%) y del método usado (28; 62%).

Discusión: Los artículos evaluados incumplen gran parte de los requisitos de calidad recogidos en la propuesta STARD, observándose una peor calidad metodológica que en otros campos diagnósticos. Los aspectos que más necesitan mejorar son los relacionados con la descripción de la selección de los pacientes y la determinación de la reproducibilidad. La investigación y el avance en nuevas tecnologías genético-moleculares requieren un mejor cumplimiento de los estándares epidemiológicos y clínicos que ya se aplican a otros campos diagnósticos.

Palabras clave: Pruebas genéticas. Metodología. Sensibilidad y especificidad.

Abstract

Introduction: The advances in genomic analysis technologies have conducted to the development of new diagnostic tests in the clinical practice. As well as it happened in other diagnostic fields, the knowledge of the main flaws affecting genetic investigation will facilitate the application of the results.

Methods: We included 44 original articles that evaluate diagnostic exactitude of genetic and molecular tests (including proteomic), published from 2002 to June 2005 in five international publications: *JAMA*, *Lancet*, *New England Journal of Medicine*, *Cancer Research* and *Clinical Cancer Research*. We examined adherence to 24 methodological criteria included in the guide STARD (Standards for Reporting of Diagnostic Accuracy).

Results: The mean number of methodological criteria satisfied was 9.8 (95% CI 8.8-10.6); the greater deficiencies were in aspects related to the description of patient's selection 9 (20%), the treatment of indeterminate results 5 (11%) and the determination of test reproducibility 6 (13%). It was observed a high fulfilment in the description of the gold standard 39 (87%) and in the methodology of the test 28 (62%).

Discussion: The methodologic quality of the evaluated articles is lower than the quality observed in other research fields. The methodologic aspects that most need improvement are those linked to the clinical information of the populations studied and the reproducibility of the tests. The research and development of new genetic-molecular technologies requires a better fulfilment of the epidemiological and clinical criteria already followed by other diagnostic fields.

Key words: Genetic techniques. Methodology. Sensitivity and specificity.

Correspondencia: Dra. Blanca Lumbreras.
Departamento de Salud Pública, Historia de la Ciencia y Ginecología. Campus de San Juan. Facultad de Medicina. Universidad Miguel Hernández.
Ctra. de Valencia, km 8,7. 03550 San Juan de Alicante. Alicante. España.
Correo electrónico: blumbreras@umh.es

Recibido: 4 de noviembre de 2005.

Aceptado: 12 de enero de 2006.

Introducción

Los avances derivados del Proyecto del Genoma Humano han supuesto el desarrollo de nuevos tipos de pruebas diagnósticas, de genética molecular o proteómica que se pueden incorporar a la práctica clínica¹⁻⁴. En España se dispone en la actualidad de pruebas basadas en el análisis de ADN para

el diagnóstico o el pronóstico de 214 enfermedades genéticas⁵, y se prevé que en los próximos años este número de pruebas se multiplique. Esto ha motivado que se hayan elaborado marcos de evaluación para la incorporación de nuevas pruebas genéticas a la cartera de servicios de los sistemas sanitarios⁶. La aparición de estas nuevas pruebas crea unas expectativas que con frecuencia no se corroboran en la práctica clínica. Esto se debe a que, en ocasiones, se publicitan resultados de investigaciones básicas sin esperar la confirmación de los resultados de las investigaciones clínicas; en otras ocasiones, la validación clínica de las pruebas carece de rigor metodológico, precisamente por desarrollarse en ámbitos de investigación básica sin el concurso de investigadores experimentados en investigación clínica y epidemiológica. En cualquier caso, las nuevas tecnologías genético-moleculares no deben introducirse para un determinado cometido clínico (cribado y consejo genético, diagnóstico, evaluación de riesgo, etc.) sin antes haber examinado tanto su validez para este propósito concreto como sus efectos en términos de salud¹.

Se ha descrito un déficit metodológico en la calidad de la investigación de otras áreas diagnósticas⁷⁻¹². En la investigación genético-molecular, en cambio, un único trabajo analizó la metodología de los estudios publicados en 4 revistas médicas internacionales y mostró que el 63% de ellos cumplía tan sólo uno de los criterios metodológicos requeridos¹³.

Para mejorar los artículos sobre investigación diagnóstica y, como consecuencia, la calidad científica de esta investigación, se ha desarrollado la iniciativa STARD (Standards for Reporting of Diagnostic Accuracy)¹⁴, cuyas recomendaciones han incorporado la mayoría de las revistas científicas. Por otra parte, Little et al¹⁵ presentaron una lista de estándares compartidos en parte por STARD, pero con algunos dirigidos únicamente a investigaciones sobre marcadores basados en genotipos, que pretendía servir de guía tanto para los autores como para los editores y revisores de artículos de genética.

Los trabajos que tratan pruebas genético-moleculares tienen una serie de especificidades, pero éstas no excluyen que se deban cumplir los criterios de calidad que se exigen en otros estudios de diagnóstico. Además, la calidad alcanzada en los estudios puede ser diferente en función de si la revista donde aparecen publicados pertenece al ámbito clínico o al diagnóstico. Para determinar si la investigación sobre diagnóstico genético-molecular alcanza la calidad exigible, se evaluaron mediante la aplicación de los criterios de la guía de STARD¹⁴ los estudios de exactitud diagnóstica genético-moleculares publicados desde el año 2002 en 4 revistas clínicas internacionales de gran importancia y en 2 revistas de la American Association for Cancer Research (AACR), caracterizadas por incluir investi-

gaciones sobre pruebas diagnósticas genético-moleculares.

Métodos

Se revisaron todos los artículos originales que estudiaban la exactitud diagnóstica de pruebas genéticas, moleculares y de proteómica desde enero de 2002 hasta junio de 2005, en 4 revistas clínicas internacionales (*JAMA*, *British Medical Journal*, *Lancet* y *New England Journal of Medicine*) y 2 revistas centradas en el campo de la investigación del cáncer pertenecientes a la American Association for Cancer Research (AACR) (*Cancer Research* y *Clinical Cancer Research*).

Selección de los artículos

Se realizó una búsqueda mediante el sistema bibliográfico informático Medline usando la estrategia de búsqueda empleada por Devillé et al¹⁶, en la que se combinaron los términos Mesh «sensitivity and specificity» con las palabras «false negative» and «accuracy». Para mejorar la sensibilidad de la búsqueda, se amplió con el término Mesh «area under the curve» y las palabras «diagnostic odds ratio» y «likelihood ratio». Tras la lectura de los resúmenes, se seleccionaron todos los artículos en los cuales se analizaba material genético humano a escala molecular y estudios de proteómica, sin limitar la búsqueda a ninguna técnica particular de laboratorio ni condición clínica.

Estándares metodológicos

Se aplicaron a cada uno de los artículos 24 de los 25 criterios pertenecientes a la guía STARD¹⁴ (se obvió el primer estándar de esta guía por tratarse del criterio de inclusión en nuestro estudio). En todos ellos se siguieron las definiciones dadas. Aunque se pueden consultar en detalle en las publicaciones de referencia, a continuación se muestran los estándares metodológicos agrupados en 4 bloques:

1. Introducción: entre los objetivos del estudio deben incluirse la estimación de la exactitud diagnóstica de una prueba o la comparación de dicha exactitud entre las pruebas o grupos participantes (1).

2. Métodos: deben aparecer los criterios de inclusión y exclusión de los participantes en el estudio (2) y el lugar donde se recogen los datos o muestras, así como una especificación del motivo de su elección (3). En el caso de que no se trate una muestra consecutiva de pacientes, se debe explicar cómo se recogie-

ron (4) y la direccionalidad del estudio (5). Sobre la prueba que se evalúa, debe definirse el patrón de referencia utilizado (6) y especificaciones técnicas que incluyan cómo y cuándo se recogieron las medidas, tanto para la prueba diagnóstica como para el patrón de referencia (7). También deben aparecer reseñadas las unidades y/o categorías de los resultados de la prueba diagnóstica y el patrón de referencia (8). Además, se deben especificar el número, el entrenamiento y la experiencia de las personas que han realizado e interpretado las pruebas y el patrón de referencia (9), y si la revisión de resultados ha sido ciega (10). Se deben describir los métodos estadísticos utilizados para calcular y comparar medidas de exactitud diagnóstica y para cuantificar el error estándar (11), así como los métodos para calcular la reproducibilidad de los tests (12).

3. Resultados: este apartado recoge características acerca del momento en que se ha realizado el estudio (13), las características clínicas y demográficas de la población en estudio (14) y el número de participantes que satisficieron los criterios de selección y que luego recibieron o no la prueba diagnóstica y/o el patrón de referencia (se puede ayudar con un diagrama de flujo) (15). También se incluye el intervalo de tiempo transcurrido entre la aplicación de la prueba diagnóstica, el patrón de referencia y cualquier otro tratamiento administrado entre ellos (16), la distribución de la severidad de la enfermedad entre los pacientes (17), una tabla con los resultados de las pruebas diagnósticas en comparación con los obtenidos del patrón de referencia (18) y la comunicación de cualquier efecto adverso producido durante el estudio (19). Por último, se debe estimar la exactitud diagnóstica y la precisión estadística (20), cómo se han tratado los resultados indeterminados (21), la estimación de la variabilidad de la exactitud diagnóstica para cada subgrupo de participantes, investigadores o centros (22) y el cálculo de la reproducibilidad de la técnica (23).

4. Discusión: en este apartado es importante que se comente la aplicabilidad clínica de los hallazgos del estudio (24).

Variabilidad de los observadores

Para calcular la variabilidad interobservador en la aplicación de los criterios, 4 de los 44 trabajos fueron seleccionados de manera aleatoria y revisados por los 2 investigadores de manera independiente. Se alcanzó un grado de acuerdo entre ellos del 86% (IC del 95%, 78-92).

La creación, las gestión de la base de datos y los análisis estadísticos de los resultados se efectuaron con el paquete estadístico SPSS (versión 12 para Windows Inc., Chicago, Illinois).

Resultados

En el período de estudio, 44 artículos cumplieron los criterios de inclusión: 12 de *Lancet*, 6 de *JAMA*, 2 de *New England Journal of Medicine*, ninguno de *BMJ*, 14 de *Cancer Research* y 10 de *Clinical Cancer Research*.

Todos los artículos revisados evaluaban pruebas centradas en el estudio del cáncer, y el de próstata 10 (22,8%) y mama 7 (15,9%) eran los más frecuentes (tabla 1). La mayoría de las pruebas diagnósticas evaluadas 37 (84%) eran genéticas, tanto mediante sistemas de micro-arrays 15 (41%) como de análisis de secuenciación genética 12 (59%), y tan sólo 7 (16%) eran de proteómica. Del total de los trabajos, 7 (16%) tenían carácter pronóstico y la mayoría, 37 (84%), diagnóstico.

La media de cumplimiento de los 24 estándares metodológicos aplicados fue de 9,8 (IC del 95%, 8,8-10,6), y ninguno de los 44 artículos evaluados siguió más de 18 criterios. No se observaron diferencias estadísticamente significativas en el cumplimiento de los criterios antes y después del año 2003 (publicación de la guía STARD¹⁴).

Al analizar los criterios de manera individual, las mayores deficiencias se advirtieron en el apartado Material y métodos (tabla 2). Apenas 11 trabajos (24%) especificaban entre sus objetivos la determinación de dicha exactitud diagnóstica. También fue baja la frecuencia (23%) con la que se especificaban los criterios de inclusión y exclusión y el lugar de procedencia de los sujetos, así como la forma de selección (20%) o si ésta se había planificado antes o después de que se realizara la prueba diagnóstica (27%). Sobre la realización de la prueba, únicamente se obtuvo un bajo cumplimiento en la descripción de las características de las personas que evalúan los tests en un trabajo (2%) y si

Tabla 1. Descripción de las enfermedades evaluadas en los trabajos analizados sobre investigación de pruebas genéticas, moleculares y de proteómica en *JAMA*, *New England Journal of Medicine*, *Lancet*, *Cancer Research* y *Clinical Cancer Research* (2002-2005)

Enfermedades	n	%
1. Cáncer de próstata	10	22,8
2. Cáncer de mama	7	15,9
3. Cáncer renal	5	11,4
4. Cáncer de ovario	4	9,1
5. Carcinoma hepatocelular	4	9,1
6. Melanoma	4	9,1
7. Cáncer de colon	4	9,1
8. Cáncer de pulmón	2	4,5
9. Neuroblastoma	2	4,5
10. Cáncer de vejiga	2	4,5

Tabla 2. Cumplimiento de los criterios relativos a «Material y métodos» por artículos sobre investigación de pruebas genéticas, moleculares y de proteómica en JAMA, New England Journal of Medicine, Lancet, Cancer Research y Clinical Cancer Research (2002-2005)

Criterios	JAMA, Lancet y New England Journal of Medicine (n = 20)	Cancer Research y Clinical Cancer Research (n = 24)	Total (n = 44)
1. Objetivos	5 (25%)	6 (25%)	11 (24%)
<i>Participantes</i>			
2. Criterios de inclusión y exclusión	8 (40%)	2 (8%)	10 (23%)
3. Recogida de pacientes	18 (90%)	22 (92%)	40 (89%)
4. Selección de la muestra	6 (30%)	3 (13%)	9 (20%)
5. Dirección del estudio	7 (35%)	5 (21%)	12 (27%)
<i>Prueba</i>			
6. Descripción del patrón de referencia	17 (85%)	22 (92%)	39 (87%)
7. Especificaciones técnicas: cómo y cuándo	7 (35%) ^a	21 (88%) ^a	28 (62%)
8. Definición de las unidades	7 (35%) ^a	20 (83%) ^a	27 (61%)
9. Personas que realizan el test	1 (5%)	0	1 (2%)
10. Revisión ciega	11 (55%) ^a	1 (4%) ^a	12 (27%)
<i>Métodos estadísticos</i>			
11. Métodos estadísticos	9 (45%)	17 (71%)	26 (58%)
12. Métodos de reproducibilidad	4 (20%)	6 (25%)	10 (23%)

^aDiferencias estadísticamente significativas ($p < 0,05$) entre los 2 grupos de revistas calculadas por la prueba de Fisher.

en la prueba la interpretación de los resultados se realizó de manera ciega, 12 artículos (27%). La descripción de los métodos para estudiar la reproducibilidad de la prueba sólo se comunicó en 10 estudios (23%), mientras que tan sólo 6 artículos (13%) mostraban los resultados obtenidos.

Respecto a la presentación de resultados (tabla 3), la mayoría definen las características clínicas y demográficas de la población en estudio 34 (77%), pero apenas se detallan los motivos por los que algunos pacientes que cumplieron los criterios de inclusión no siguieron posteriormente en el estudio 11 (25%). Sólo 12 estudios (27%) investigaron si la sensibilidad y la especificidad diagnósticas variaban de acuerdo con subgrupos clínicos relevantes. La comunicación y el análisis de los resultados indeterminados o poco precisos fue muy infrecuente 5 (11%).

Si comparamos el cumplimiento de los criterios entre las revistas clínicas y las dedicadas a la investigación del cáncer (tablas 2 y 3), podemos comprobar que hay diferencias estadísticamente significativas en determinados estándares. Las revistas relacionadas con el cáncer cumplieron mejor que las publicaciones clínicas la definición de las especificaciones técnicas de la prueba, con 21 (88%) y 7 (35%), respectivamente, así como la especificación de las unidades empleadas en la prueba, con 20 (83%) respecto a 7 (35%). También expresaron con mayor frecuencia la correspondencia de los resultados del patrón de referencia con la prueba evaluada (22 [92%] frente a 8 [40%]), y el estudio de la variación de la exactitud según subgrupos clínicos rele-

vantes (10 [42%] frente a 2 [10%]). Sin embargo, se comprobó un mayor cumplimiento de las revistas clínicas en comparación con las de cáncer en los estándares relacionados con la revisión ciega de los resultados (11 [55%] y 1 [4%], respectivamente) y en la presentación de las estimaciones de exactitud y su precisión estadística (12 [60%] frente a 6 [25%]).

Discusión

Los artículos sobre pruebas de diagnóstico genético-molecular publicados recientemente (2002-2005) en las mejores revistas de medicina clínica internacionales y en 2 publicaciones importantes de investigación en cáncer incumplen gran parte de los requisitos de calidad recogidos en la propuesta STARD¹⁴. Este incumplimiento puede ser reflejo de graves limitaciones en la metodología de la investigación, de descuido en la preparación de los artículos y su proceso editorial, o de ambas cosas. Si los comparamos con la calidad observada en otros campos diagnósticos⁹⁻¹², comprobamos que los estudios de genética, proteómica y moleculares muestran una calidad metodológica inferior, mientras que, por ejemplo, en el campo del laboratorio, y debido fundamentalmente a los esfuerzos realizados por distintos autores y grupos editoriales, se observa paulatinamente un avance en su calidad¹¹.

La calidad metodológica encontrada tanto en las revistas clínicas como en las dedicadas a la investigación

Tabla 3. Cumplimiento de los criterios relativos a los resultados y discusión de los artículos sobre pruebas diagnósticas genéticas, moleculares y de proteómica en JAMA, New England Journal of Medicine, Lancet, Cancer Research y Clinical Cancer Research (2002-2005)

Criterios	JAMA, Lancet y New England Journal of Medicine (n = 20)	Cancer Research y Clinical Cancer Research (n = 24)	Total (n = 44)
<i>Participantes</i>			
13. Fechas de realización del estudio	12 (60%)	6 (25%)	18 (41%)
14. Descripción de las características de los sujetos	15 (75%)	19 (79%)	34 (77%)
15. Presentación del diagrama de flujo	8 (40%)	3 (13%)	11 (25%)
<i>Resultados de la prueba</i>			
16. Tiempo entre la práctica de la prueba y el patrón de referencia	4 (20%)	11 (46%)	15 (34%)
17. Distribución de la enfermedad	16 (80%)	21 (86%)	37 (84%)
18. Resultados de la prueba y patrón de referencia	8 (40%) ^a	22 (92%) ^a	30 (68%)
19. Comunicación de efectos adversos	0	0	0
<i>Estimaciones</i>			
20. Presentación de las estimaciones de exactitud y su precisión estadística	12 (60%) ^a	6 (25%) ^a	18 (41%)
21. Comunicación de los resultados indeterminados	3 (15%)	2 (8%)	5 (11%)
22. Estudios de la variación de la exactitud según subgrupos clínicos relevantes	2 (10%) ^a	10 (42%) ^a	12 (27%)
23. Determinación de la reproducibilidad	3 (15%)	3 (13%)	6 (13%)
<i>Discusión</i>			
24. Valoración de la aplicabilidad clínica	9 (45%)	9 (42%)	18 (42%)

^aDiferencias estadísticamente significativas ($p < 0,05$) calculadas mediante el test de la χ^2 entre los 2 grupos de revistas calculadas por la prueba de Fisher.

del cáncer es parecida, aunque con alguna especificidad. Los aspectos más analíticos, como la especificación de las características de la técnica, están más cuidados en las revistas de cáncer, mientras que otros más relacionados con la manera de llevar a cabo el estudio, como la revisión ciega de los resultados obtenidos, se cumplen más en las revistas clínicas.

El único antecedente en el campo del diagnóstico genético-molecular de este tipo de investigación fue una revisión realizada en 1995 en las mismas 4 revistas clínicas analizadas en este trabajo¹³. A pesar de que no se utilizaron los mismos criterios metodológicos, los empleados entonces están incluidos en la guía STARD y, por lo tanto, son comparables. Al igual que en esta revisión, ningún trabajo cumplió un número aceptable de estándares metodológicos, y cabe resaltar lo inusual del cálculo de la reproducibilidad de las pruebas genéticas y la ausencia de revisión ciega en la interpretación de los resultados. Parece, por lo tanto, que los estudios diagnósticos genético-moleculares son más impermeables a los avances clínico-epidemiológicos.

Nuestra intención de facilitar la comparación con trabajos previos nos llevó a la aplicación de los mismos criterios restrictivos de selección antes empleados^{7,9-12} y, en consecuencia, a una muestra reducida de artículos que limita el alcance de los resultados. Sin embargo, la inclusión de todos los artículos originales sobre el campo del diagnóstico genético y molecular publicados en las mejores revistas internacionales de medicina clí-

nica y en las 2 publicaciones de la American Association Cancer Research (por contener una parte importante de los trabajos publicados sobre la cuestión abordada) ofrece una adecuada perspectiva sobre la calidad de investigación y publicación de este tipo de estudios. La guía empleada para la evaluación de los artículos (STARD) es muy reciente, pero se ha difundido rápidamente en la mayoría de revistas que publican evaluaciones diagnósticas; de hecho, dos de ellas (*JAMA* y *Lancet*) las han incluido entre las normas para los autores^{17,18}, pero no *New England Journal of Medicine* y las 2 revistas de cáncer. Se podría argumentar que hay escasa experiencia en su aplicación y es cierto que puede haber dudas sobre la aplicabilidad de algunos de sus criterios por falta de indicaciones precisas. No obstante, la experiencia disponible muestra que hay una elevada concordancia entre observadores al aplicarla^{11,12}, como también la ha habido en la muestra realizada en el presente estudio. Además, muchos de los criterios de STARD ya han sido previamente aplicados en anteriores guías metodológicas⁷⁻¹², lo que ha facilitado su empleo. Los estudios de diagnóstico en genómica, proteómica y campos relacionados tienen unas especificidades en su validación¹⁹ que no recoge la guía STARD y, aunque se han originado diversas iniciativas para la elaboración de una guía de recomendaciones apta para la validación de estas pruebas¹⁵, todavía no se ha esbozado ninguna definitiva. Este hecho supone una limitación en la medida en que no se han eva-

luado algunos problemas intrínsecos, aunque no exclusivos de este campo de la investigación. Un ejemplo es el *overfitting*, que se produce cuando, por ejemplo, se propone un perfil proteómico seleccionado de un conjunto más amplio tras innumerables ensayos de discriminación y que afecta a la validez externa. Este problema se controla mediante la determinación de la reproducibilidad del patrón discriminatorio observado en muestras independientes de la original. Si el patrón de discriminación no puede ser reproducido en otras muestras, es probable que haya *overfitting*²⁰.

Otra limitación que se debe tener en cuenta es el período de estudio, ya que los criterios STARD aparecieron publicados a principios del año 2003¹⁴, y este estudio recoge artículos desde 2002 hasta 2005. Ciertamente, es pronto para que la iniciativa tenga efecto, al menos en las revistas que la han incluido en las instrucciones para sus autores, y nuestro estudio no tiene potencia estadística para establecer comparaciones. Sin embargo, el hecho de que no estuviera publicada no explica la baja calidad observada, ya que muchos de sus criterios son requisitos imprescindibles en la mayoría de estudios observacionales ya recogidos parcialmente –los más relevantes– en otras guías y publicaciones anteriores.

No todos los requisitos tienen, lógicamente, la misma importancia en cuanto a sus efectos en la validez de los estudios evaluados. Teniendo en cuenta las principales características que debe reunir una prueba genética para su correcta utilización²¹, es importante que resaltemos los aspectos clave para la validez que se han mostrado como más deficientes. En primer lugar, se encuentran las características analíticas de la prueba evaluada ya que, debido al gran desarrollo de pruebas genéticas que se está produciendo, para una misma prueba puede haber distintas técnicas y habrá que conocer cuál de ellas muestra unos valores más aceptables de precisión diagnóstica. Por ello, destaca el escaso cálculo de la reproducibilidad de la técnica en las investigaciones evaluadas. En revisiones anteriores, este concepto ya se consideraba imprescindible⁷⁻¹² y, sin embargo, en el campo de la genética o la proteómica, donde el valor de las determinaciones también depende de su capacidad de proporcionar el mismo resultado cuando se aplica a los mismos pacientes en igualdad de condiciones¹⁹, se ha descuidado con excesiva frecuencia.

En segundo lugar, otra deficiencia metodológica clave y que limita la aplicabilidad de los resultados es la escasa descripción de los criterios de inclusión y exclusión de los sujetos participantes en el estudio, ya que éstos apenas aparecen en menos de la mitad de los trabajos evaluados. Esta falta de información impide considerar en detalle la aplicabilidad clínica de la prueba evaluada, que es un elemento indispensable en este tipo de investigación²¹.

En resumen, la proyección futura en el campo del diagnóstico y la detección precoz de las nuevas tecnologías genético-moleculares se encuentra muy limitada por la falta de rigor en la investigación sobre estas pruebas. Estos problemas están en el origen de las reiteradas falsas expectativas que se crean alrededor de estas pruebas diagnósticas. Si hubiese una relación efectiva interdisciplinaria (básicos, clínicos y epidemiólogos) que facilitase el desarrollo riguroso de pruebas diagnósticas genético-moleculares, se evitaría mucha investigación innecesaria. Además, el cumplimiento con los estándares metodológicos permitiría que las pruebas genético-moleculares y de proteómica se convirtiesen en herramientas muy útiles en la investigación clínica y de servicios sanitarios, especialmente en el campo del cáncer²².

Agradecimientos a subvención

Este trabajo ha sido financiado gracias a ISCIII, Red de Centros RCEP C03/09 y al Ministerio de Sanidad y Consumo, Instituto de Salud Carlos III (Ayudas para contratos post Formación Sanitaria Especializada).

Bibliografía

- Schwartz MK. Genetic testing and the clinical laboratory improvement amendments of 1988: present and future. *Clin Chem*. 1999;45:739-45.
- Higashi MK, Veenstra DL. Managed care in the genomics era: assessing the cost effectiveness of genetic tests. *Am J Manag Care*. 2003;9:493-500.
- Milunsky A. Commercialization of clinical genetic laboratory services: in whose best interest? *Obstet Gynecol*. 1993;81:627-9.
- Korf BR. Advances in molecular diagnosis. *Curr Opin Obstet Gynecol*. 1996;8:130-4.
- Rueda J, Briones E. Servicios de diagnóstico genético para las enfermedades hereditarias en España. EUR 20516 EN. Sevilla: Institute for Prospective Technological Studies; 2002.
- Márquez S, Briones E. Marco para la evaluación de las pruebas genéticas en el Sistema Sanitario Público de Andalucía. Informe 2/2005.
- 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.
- Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, Van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*. 1999;282:1061-6.
- Ramos Rincón JM, Hernández Aguado I. Investigación en pruebas diagnósticas en Medicina Clínica. Una evaluación de la metodología. *Med Clin (Barc)*. 1998;111:129-34.
- Ramos JM, Hernández I. Métodos para evaluar pruebas diagnósticas en enfermedades infecciosas y microbiología clínica. *Enferm Infecc Microbiol Clin*. 1998;16:179-84.
- Lumbreras-Lacarra B, Ramos-Rincón JM, Hernández-Aguado I. Methodology in diagnostic laboratory test research in clinical chemistry and clinical chemistry and laboratory medicine. *Clin Chem*. 2004;50:530-6.

12. Lumbreras Lacarra B, Ramos Rincón JM, Hernández Aguado I. Evaluación de la investigación metodológica en pruebas diagnósticas de laboratorio en Revista Clínica Española y Medicina Clínica. Rev Clin Esp. 2004;204:472-6.
13. Bogardus ST Jr, Concato J, Feinstein AR. Clinical epidemiological quality in molecular genetic research: the need for methodological standards. JAMA. 1999;281:1919-26.
14. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Standards for reporting of diagnostic accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. Clin Chem. 2003;49:1-6.
15. Little J, Bradley L, Bray MS, Clyne M, Dorman J, Ellsworth DL, et al. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. Am J Epidemiol. 2002;156:300-10.
16. Deville WL, Bezemer PD, Bouter LM. Publications on diagnostic test evaluation in family medicine journals: an optimal search strategy. J Clin Epidemiol. 2000;53:65-9.
17. Rennie D. Improving reports of studies of diagnostic tests: the STARD initiative. JAMA. 2003;289:89-90.
18. Bossuyt PM, Reitsma JB. Standards for Reporting of Diagnostic Accuracy. The STARD initiative. Lancet. 2003;361:71.
19. Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. Nat Rev Cancer. 2004;4:309-14.
20. Ransohoff DF. Lessons from controversy: ovarian cancer screening and serum proteomics. J Natl Cancer Inst. 2005;97:315-9.
21. Burke W. Genetic testing. N Engl J Med. 2002;347:1867-75.
22. Porta M, Fernández E, Alguacil J. Semiology, proteomics and the early detection of symptomatic cancer. J Clin Epidemiol. 2003;56:815-9.