

Estudios de precisión diagnóstica (STARD) y pronóstica (REMARK)

Douglas G. Altman^a y Patrick M.M. Bossuyt^b

^aCentre for Statistics in Medicine. Wolfson College. Oxford. Reino Unido.

^bDepartment of Clinical Epidemiology & Biostatistics. Academic Medical Center. University of Amsterdam. Amsterdam. Países Bajos.

Existen deficiencias manifiestas en la comunicación de los estudios de investigación, que abarcan todas las especialidades y todos los diseños. Aunque los ensayos clínicos efectuados con asignación aleatoria han sido los que han recibido mayor atención, han surgido preocupaciones similares en otros tipos de estudios, como los de carácter diagnóstico y epidemiológico.

Si el artículo publicado describe con suficiente detalle la metodología aplicada en cada fase, el lector recibe suficiente información para decidir por sí mismo sobre la solidez de los resultados obtenidos. El fundamento científico del desarrollo de directrices sobre la comunicación de los estudios de investigación se basa en esta sencilla idea. Las recomendaciones para las directrices sobre los informes de investigaciones han evolucionado a lo largo de los años, a través de la preparación de una serie de directrices de comunicación que se iniciaron con los proyectos CONSORT y QUOROM en la década de los noventa. En este artículo se describen las iniciativas desarrolladas para el establecimiento de directrices de comunicación sobre los estudios de precisión de pruebas diagnósticas (STARD) y del valor pronóstico de los marcadores tumorales (REMARK).

Palabras clave: Comunicación de la investigación. Directrices. Diagnóstico. Pronóstico. STARD. REMARK.

Diagnostic (STARD) and prognostic (REMARK) studies

Deficiencies in how research studies are reported are both well-documented and widespread across all medical specialties and study designs. Although randomised trials have received the most attention in this regard, similar concerns have been expressed about reporting of other types of research including diagnostic and epidemiological studies.

If a journal article describes in enough detail what was done at each stage of a study, readers will have enough information to allow them to decide on the merits of the results for themselves. From this simple idea comes the scientific rationale of developing guidelines on how to report research.

Recommended processes to produce reporting guidelines have evolved over several years during the preparation of a sequence of reporting guidelines starting with CONSORT and QUOROM in the 1990s. We describe initiatives to develop reporting guidelines for diagnostic accuracy studies (STARD) and tumour marker prognostic studies (REMARK).

Key words: Reporting research. Guidelines. Diagnosis. Prognosis. STARD. REMARK.

Introducción

La comunicación precisa y transparente de la investigación se ha convertido en una cuestión cada vez más relevante a medida que se acumulan pruebas de que muchos de los estudios de investigación publicados no ofrecen información clave respecto a su metodología y sus resultados. A su vez,

Douglas G. Altman ha recibido una ayuda económica de Cancer Research UK.

Correspondencia: Dr. Douglas G. Altman
Centre for Statistics in Medicine. Wolfson College.
Linton Road, Oxford OX2 6UD, Reino Unido.
Correo electrónico: doug.altman@cancer.org.uk

esta preocupación ha generado una serie de iniciativas para definir los aspectos que deben recogerse en la publicación de un estudio.

Hasta hace relativamente poco, las iniciativas para el establecimiento de recomendaciones respecto a la comunicación de los estudios de investigación se han referido, casi exclusivamente, a los ensayos clínicos efectuados con asignación aleatoria y control, sobre todo la Declaración CONSORT¹. Sin embargo, la comunicación de los resultados de cualquier otro tipo de investigación médica presenta deficiencias similares, de manera que varios grupos de estudio han ampliado la idea del desarrollo de directrices de comunicación para abarcar también a estos otros tipos de estudio. En este artículo se describen las distintas iniciativas para la elaboración de directrices para la comunicación de los estudios relativos a la precisión de las pruebas diagnósticas (STARD) y a los marcadores pronósticos (REMARK).

Estudios relativos a la precisión de las pruebas diagnósticas

Contexto

Durante las últimas décadas, el número de pruebas médicas se ha incrementado de manera rápida, tanto en lo que se refiere a las nuevas pruebas como a la tecnología de las pruebas ya introducidas, que mejoran de manera continua. Las pruebas médicas no sólo se utilizan para facilitar el diagnóstico, sino también para la toma de decisiones respecto al tratamiento, a la respuesta a éste y a la detección de posibles trastornos subclínicos, así como para la determinación de la salud futura del paciente^{2,3}.

Al igual que ocurre con las nuevas tecnologías médicas, las nuevas pruebas diagnósticas deben evaluarse con detalle antes de su introducción en la práctica clínica. El proceso riguroso de evaluación de las pruebas diagnósticas antes de su introducción no sólo reduce el número de consecuencias clínicas no deseadas, secundarias a las estimaciones erróneas de la precisión diagnóstica de la prueba, sino que también limita los costes sanitarios, ya que evita la realización de pruebas innecesarias.

Por desgracia, la evaluación de la precisión diagnóstica de las pruebas médicas va con retraso respecto a la evaluación de los tratamientos^{2,3}. Los resultados exagerados y sesgados que se obtienen en estudios diagnósticos mal diseñados y pobemente comunicados dan lugar a una difusión anticipada del valor de las pruebas diagnósticas y hacen que los clínicos tomen decisiones terapéuticas incorrectas. Además, es necesario tener en cuenta que los fondos de financiación de la investigación no suelen cubrir la realización de estudios fundamentados en síntomas o pruebas diagnósticas.

En la evaluación de las pruebas médicas, tanto de las nuevas como de las ya existentes, un paso esencial es el establecimiento de la precisión de la prueba diagnóstica. En los

estudios de precisión diagnóstica, se comparan los resultados de una o más pruebas con los obtenidos mediante la prueba diagnóstica de referencia sobre el mismo grupo de pacientes en los que se considera que hay un proceso patológico concreto. El término *prueba* se refiere a cualquier método utilizado para obtener información adicional sobre el estado de salud de un paciente. Incluye la información conseguida a través de la historia clínica, la exploración física, las pruebas analíticas, de imagen, funcionales y el estudio anatopatológico. La enfermedad objetivo o el proceso patológico de interés son un trastorno concreto, una fase de la enfermedad o cualquier otro proceso patológico identificable que obligue a acciones clínicas, como la realización de nuevas pruebas diagnósticas o el inicio, la modificación o la interrupción de un tratamiento. En este contexto, la prueba diagnóstica de referencia se considera el mejor método para establecer la presencia o la ausencia de un proceso patológico. La prueba de referencia puede ser un método único o una combinación de métodos, y permite establecer la presencia de la enfermedad objetivo. Puede ser una prueba analítica, un estudio de imagen o un estudio anatopatológico, pero también un método concreto de seguimiento clínico de los pacientes. El término *precisión* se refiere al grado de concordancia entre la información ofrecida por la prueba diagnóstica evaluada (denominada prueba índice) y la ofrecida por la prueba de referencia. La precisión diagnóstica se puede expresar de muchas maneras, como los niveles de sensibilidad y especificidad, los cocientes de probabilidad, el cociente de posibilidades diagnósticas y el área bajo la curva de eficacia diagnóstica (ROC, *receiver operator characteristic*).

Hay varios factores que amenazan la validez interna y externa de los resultados obtenidos en un estudio de precisión diagnóstica de una prueba. Algunos de estos factores se relacionan con el diseño de este tipo de estudios, otros con la forma de selección de los pacientes, con el método de realización de las pruebas o con el análisis de los datos obtenidos⁴. Un estudio basado en varios metaanálisis ha demostrado que la apreciación excesivamente optimista de la precisión de las pruebas diagnósticas puede deberse a diversas deficiencias en el diseño de su estudio⁵.

El número de estudios relativos a la precisión de las pruebas diagnósticas publicados es cada vez mayor, pero la calidad es, en general, escasa. En una encuesta relativa a los estudios de precisión diagnóstica publicada entre 1978 y 1993 en cuatro importantes revistas de medicina se demostró que la calidad metodológica de esos estudios era, como mucho, mediocre⁶. En encuestas más recientes se han observado deficiencias similares⁷.

En esta revisión también se demostró que la publicación de los estudios no contenía información esencial sobre puntos clave, como su diseño, sus métodos de realización y los análisis de sus resultados. La ausencia de información clave respecto al diseño y a la realización de los estudios diagnósticos ha sido confirmada por los autores de diversas revisiones sistemáticas (metaanálisis). Éste es un problema importante, ya que la comunicación completa y precisa permitiría que el lector pudiera detectar los posibles sesgos que aquejan a un estudio, así como juzgar la posibilidad de generalización y aplicación de los resultados.

Desarrollo de directrices para la comunicación de los estudios relativos a la precisión de las pruebas diagnósticas (Declaración STARD)

En la reunión de la Colaboración Cochrane que tuvo lugar en 1999 en Roma, el grupo de trabajo sobre métodos diagnósticos y pruebas de detección (*Screening Test Methods*

Working Group) de la propia Colaboración Cochrane expuso la escasa calidad metodológica y las deficiencias en la comunicación de estudios de este tipo. El grupo consideró que el primer paso para corregir estos problemas era mejorar la calidad de la comunicación de los estudios diagnósticos. Entonces se constituyeron los Estándares para el Informe de Estudios de Precisión Diagnóstica (STARD, *Standards for Reporting of Diagnostic Accuracy*), cuyo objetivo fue mejorar la calidad de la comunicación de los estudios de precisión de pruebas diagnósticas. Tras el éxito de la Declaración CONSORT, el grupo STARD se propuso como objetivo el desarrollo de una lista de puntos que se deberían abordar en cualquier estudio relativo a la precisión de las pruebas diagnósticas.

El Comité de dirección STARD realizó una búsqueda detallada de la bibliografía y elaboró una lista de 75 puntos posibles. Esta búsqueda se llevó a cabo en las bases de datos Medline, Embase y BIOSIS, y en la base de datos metodológica de la Colaboración Cochrane, hasta julio de 2000. Además, los miembros del comité de dirección STARD examinaron las listas de bibliografía de los artículos seleccionados, efectuaron búsquedas en sus archivos bibliográficos personales y establecieron contactos con otros expertos en el campo de la investigación diagnóstica. Revisaron todas las publicaciones relevantes y efectuaron una ampliación de la lista de posibles puntos.

Hay dos consideraciones generales que son útiles para determinar el contenido y el formato de la lista. En primer lugar, el grupo STARD considera que el uso de una sola lista tiene más posibilidades de ser conocida y aceptada por autores, revisores y editores de revistas médicas que el uso de varias listas distintas. A pesar de que la evaluación de una prueba de imagen es diferente de la evaluación de una prueba analítica, estas diferencias pueden ser más de forma que de fondo. La segunda consideración se refiere al desarrollo de una lista dirigida específicamente hacia los estudios sobre la precisión de las pruebas diagnósticas. El grupo STARD no incluyó aspectos generales en la comunicación de los resultados de los estudios de investigación, como las recomendaciones del documento de Requisitos de Uniformidad de los Originales Remitidos a las Revistas de Biomedicina (*Uniform Requirements for Manuscripts submitted to Biomedical Journals*).

Posteriormente, el comité de dirección STARD, en una reunión de consenso de 2 días de duración (16 y 17 de septiembre de 2001) en Amsterdam, y a la que fueron invitados expertos investigadores, editores, especialistas en metodología y organizaciones profesionales, evaluaron todos los temas propuestos y elaboraron una lista final de 25 puntos.

Además, el grupo STARD desarrolló un esfuerzo considerable para el establecimiento de un diagrama de flujo relativo a los estudios de precisión de las pruebas diagnósticas. Este diagrama de flujo informa sobre el método de selección de los pacientes, el orden de realización de las pruebas y el número de pacientes que deben evaluarse mediante la prueba índice y la prueba de referencia. El diagrama de flujo comunica de manera transparente información clave respecto al diseño de un estudio y el flujo de los participantes. Diversos grupos de usuarios potenciales evaluaron la primera versión de la lista y del diagrama de flujo. Esta lista se ofreció en el sitio web de CONSORT (www.consort-statement.org), con una solicitud de comentarios. El grupo STARD recibió comentarios y observaciones muy útiles durante las diversas fases de evaluación, y elaboró una lista final de puntos de una sola página. La tabla 1 muestra la lista de directrices STARD, y la figura 1, un diagrama de flujo prototípico relativo a un estudio de precisión diagnóstica.

TABLA 1

Lista de elementos STARD para la publicación de estudios relativos a la precisión diagnóstica

	Punto n°	Descriptor	Informado en página número
Título/Resumen/Palabras clave	1	Identifica el artículo como un estudio de precisión diagnóstica (epígrafe MeSH recomendado «sensibilidad y especificidad»)	
Introducción	2	Especifica los objetivos del estudio, como estimación de la precisión diagnóstica o comparación de la precisión entre distintos grupos	
Métodos			
Participantes	3	Población estudiada; criterios de inclusión y exclusión, contexto y centros en que se obtuvieron los datos	
	4	Proceso de selección de los participantes: ¿el proceso de selección estuvo fundamentado en los síntomas iniciales, en los resultados obtenidos en pruebas previas o en el hecho de que los participantes se hubiesen evaluado mediante la prueba índice o estándar de referencia?	
	5	Selección de los participantes: ¿la población de pacientes estudiada constituyó una serie consecutiva según los criterios de selección de los puntos 3 y 4? En caso negativo, especifique el método de selección empleado	
	6	Recogida de los datos: ¿se había planificado antes (estudio prospectivo) o bien después (retrospectivo) de la realización de las pruebas índice y de referencia?	
Realización de la prueba	7	Prueba estándar de referencia y su fundamento	
	8	Especificaciones de carácter técnico sobre materiales y métodos para la realización de las pruebas, incluyendo tipo, momento y referencias bibliográficas	
	9	Definición y fundamento de las unidades, valores umbral y categorías de los resultados de las pruebas índice y de referencia	
	10	Número y formación de las personas que realizan las pruebas. Coste económico	
	11	Especifique si los investigadores que interpretan todas las pruebas permanecen cegados. Describa toda información clínica relevante	
Métodos estadísticos	12	Métodos para el cálculo o la comparación de la precisión diagnóstica y para cuantificar la incertidumbre (p. ej., intervalos de confianza del 95%)	
Resultados			
Participantes	13	Si procede, métodos para el cálculo de la reproducibilidad y fiabilidad de la prueba	
	14	Fechas de realización del estudio, incluyendo las de comienzo y de finalización de la recogida de datos	
	15	Características clínicas y demográficas de los participantes en el estudio (p. ej., edad, sexo, espectro de síntomas iniciales, comorbilidad, tratamientos actuales, centros)	
	16	Número de pacientes que cumplen los criterios de participación en el estudio y que se evalúan o no mediante las pruebas índice y de referencia; descripción de las razones por las que los pacientes no se evaluaron mediante alguna de estas pruebas (se recomienda el uso de un diagrama de flujo)	
Resultados de la prueba	17	Intervalos de tiempo transcurridos entre la realización de las pruebas índice y estándar, así como descripción de cualquier tratamiento utilizado entre ambas	
	18	Descripción de la gravedad de la enfermedad (definir los criterios) en los pacientes que la presentan y descripción de otros diagnósticos en aquellos que no la presentan	
	19	Tabla de resultados obtenidos con la prueba índice (incluyendo los indeterminados y los no obtenidos) según los resultados de la prueba estándar. Para variables continuas, la distribución de los resultados según los valores del estándar de referencia	
	20	Todo evento adverso observado durante la realización de las pruebas	
Estimaciones	21	Estimaciones de la precisión diagnóstica y de la incertidumbre estadística (p. ej., intervalos de confianza del 95%)	
	22	Métodos para considerar los resultados extremos o no determinados y los datos ausentes	
	23	Si procede, estimación de la heterogeneidad de la precisión diagnóstica entre grupos de participantes, de investigadores o de centros	
	24	Si procede, estimación de la reproducibilidad o fiabilidad de la prueba	
Discusión	25	Discusión de la aplicabilidad clínica de los resultados del estudio	

Publicación

La Declaración STARD se publicó en los primeros ejemplares de 2003 de siete importantes revistas médicas generales y de especialidad, como *Annals of Internal Medicine*, *Radiology*, *The British Medical Journal* y *Clinical Chemistry*⁸, y posteriormente en algunas otras revistas biomédicas. La declaración y la lista se acompañaban de un documento explicativo⁹, en el que se exponían la significación y el fundamento de cada uno de los puntos, y se revisaba brevemente

la evidencia respecto a estos. Los documentos también se recogieron en varios sitios web, como el de *Clinical Chemistry* y el del grupo CONSORT.

La publicación de los documentos STARD se acompañó de varios editoriales en los que se recomendaba su aceptación por parte de los editores de revistas y de los comités de revisión. En un estudio de evaluación que ha finalizado recientemente y que se ha referido a los estudios sobre la precisión de las pruebas diagnósticas publicados en 2000 (antes de la

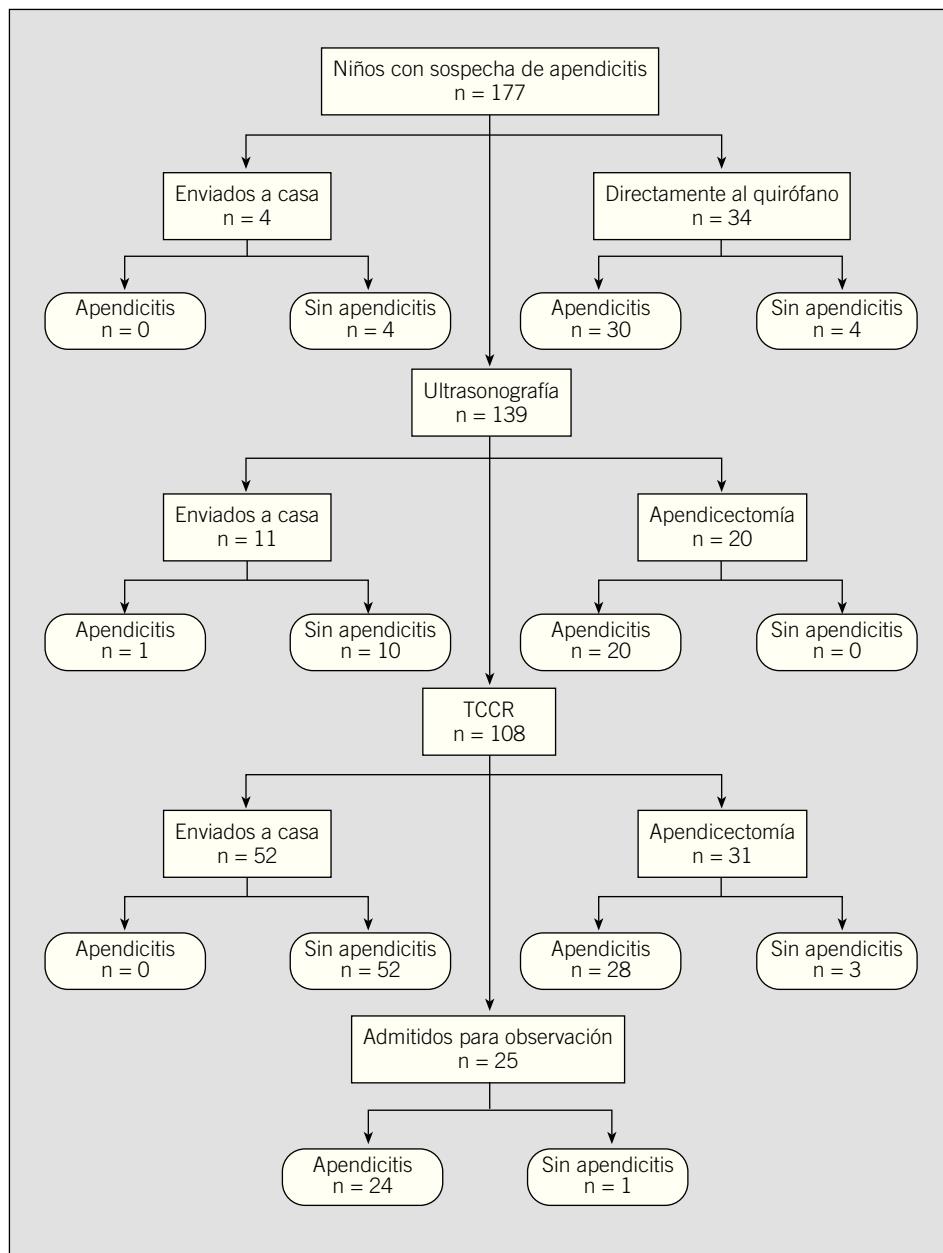


Fig. 1. Diagrama de flujo prototípico de un estudio de precisión diagnóstica. TCCR: tomografía computarizada con contraste renal.

Declaración STARD) y en 2004 (después de la Declaración STARD) en revistas de gran impacto, se demuestra que la calidad de la comunicación en los artículos relativos a la precisión diagnóstica ha mejorado tras la publicación de la Declaración STARD, pero queda todavía camino por recorrer¹⁰. Como promedio, en los artículos publicados en 2004 se han recogido 14 de los 25 puntos esenciales de la lista. El grupo STARD tiene planificada la actualización de la Declaración STARD cuando aparezcan nuevas evidencias sobre el origen de los sesgos o de la variabilidad. Si las revistas médicas, los autores, los editores y los revisores adoptaran con mayor interés tanto la lista de puntos como el diagrama de flujo, mejoraría la calidad de la publicación de los estudios relativos a la precisión diagnóstica de las pruebas para beneficio tanto de los clínicos como de los investigadores, los revisores, las revistas y la sociedad en general.

Estudios sobre el valor pronóstico de los marcadores tumorales

Contexto

A lo largo de muchos años, la publicación de miles de estudios sobre marcadores tumorales en oncología ha culminado con la demostración de la utilidad clínica real para sólo unos pocos¹¹. Los estudios iniciales sobre un marcador dan lugar, a menudo, a resultados muy prometedores, pero en los estudios siguientes sobre ese marcador o sobre otros marcadores relacionados se obtienen conclusiones contradictorias o claramente incongruentes con los prometedores resultados iniciales. Se han citado muchos problemas para explicar esta discrepancia, como las diferencias metodológicas, la baja calidad del diseño de los estudios, la variabilidad y la falta de definición de los criterios de participación de los pacientes, y la inexistencia de posibilidades de reproducibilidad del estudio

efectuado. También puede haber un sesgo o una variabilidad importantes de manera inherente al uso de muchos métodos de determinación distintos¹². Los análisis estadísticos inadecuados son habituales, como los estudios en los que la potencia estadística del método utilizado es insuficiente y en los que se obtienen resultados excesivamente optimistas respecto a la magnitud del efecto o el nivel de significación a consecuencia de la realización de múltiples pruebas, análisis de subgrupos u optimización de los valores umbral¹³⁻¹⁶.

Los lectores de las publicaciones de estos estudios necesitan saber con precisión qué es lo que se ha realizado, con objeto de que puedan juzgar la fiabilidad de los resultados. Por desgracia, los estudios sobre marcadores tumorales presentan una baja calidad de publicación; muchos de los artículos publicados carecen de la información suficiente como para que el lector pueda valorar de manera adecuada los métodos utilizados o la posibilidad de generalización de sus resultados. Las deficiencias en la publicación de los estudios de carácter pronóstico (así como su heterogeneidad metodológica) quedan en evidencia a través de las revisiones sistemáticas de la bibliografía sobre marcadores o tumores malignos concretos¹⁷⁻²². Incluso la presentación de los resultados reales tiene generalmente una calidad muy baja¹⁸.

Una recomendación importante de la NCI-EORTC First International Meeting on Cancer Diagnostics (From Discovery to Clinical Practice: Diagnostic Innovation, Implementation, and Evaluation), celebrada en Nyborg, Dinamarca, en julio de 2000, fue el desarrollo de directrices para la publicación de los estudios relativos a los marcadores tumorales. En esa reunión se consideró que algunas de las barreras principales para el progreso en esta área eran la baja calidad del diseño y el análisis de los estudios, la variabilidad en las pruebas utilizadas y la comunicación deficiente de los estudios. El subcomité de estadística quedó encargado de abordar los problemas estadísticos relativos a las deficiencias en el diseño y en el análisis de los estudios sobre marcadores tumorales pronósticos. Se acordó que la prioridad debía ser la elaboración de un documento de directrices para la publicación.

Desarrollo de directrices para la comunicación de los resultados obtenidos en estudios sobre el valor pronóstico de los marcadores tumorales (REMARK)

Los 9 miembros del comité, pertenecientes a Estados Unidos y a Europa, constituyan un grupo heterogéneo de clínicos, especialistas en estadística y científicos de laboratorio. Algunos de ellos realizaban editoriales para revistas de oncología, pero ninguno era editor.

Las ideas iniciales respecto a los puntos clave del documento de directrices se extrajeron a partir de la bibliografía sobre la evidencia empírica relativa a las deficiencias en la publicación o a los métodos de análisis problemáticos, las sugerencias previas relacionadas con la publicación de estos estudios y las diversas publicaciones de tipo educativo. También se generaron ideas a través de la revisión de las directrices de publicación similares elaboradas respecto a otros tipos de estudios de investigación médica, como los documentos CONSORT¹, QUOROM²³ y STARD⁹. Tres de los miembros de este grupo escribieron un primer borrador que sirvió como base para la discusión de todo el grupo. Los comentarios al borrador se realizaron a través de una videoconferencia de todo el grupo, y mediante numerosos intercambios por correo electrónico, pero no se llevó a cabo ninguna reunión presencial. Tras otras revisiones efectuadas por el grupo y por expertos externos, la versión de febrero de 2003 se remitió a todo el grupo para su comentario final.

El hecho de que hayan transcurrido dos años antes de su publicación puede atribuirse a varios factores. El principal, la intención frustrada de desarrollar un documento con una arquitectura similar a la de los documentos de las declaraciones CONSORT y STARD^{9,24}. Finalmente, se decidió la publicación de las recomendaciones solas, antes de que ésta se retrasara todavía más. El documento explicativo va a incluir las justificaciones específicas de la necesidad de cada uno de los puntos contenidos en las recomendaciones. Actualmente, el documento conjunto está en fase de desarrollo y se espera que sea publicado en 2006.

La tabla 2 muestra la lista de directrices REMARK, relativa a las recomendaciones para la publicación de los estudios sobre marcadores tumorales. Los puntos específicos aparecen agrupados bajo epígrafes (Introducción, Materiales y métodos, Resultados y Discusión) que reflejan las secciones de los artículos científicos publicados.

Tal como se puede observar en el elemento 12, el diagrama debe ser útil para que sea posible determinar el número de participantes en cada una de las distintas fases de un estudio. Como mínimo, este diagrama debe incluir el número de pacientes de la muestra original, el número de pacientes que permanecen en el estudio tras la aplicación de los criterios de exclusión y el número de pacientes evaluados a través de los análisis uni y multivariados.

Las directrices REMARK están centradas, principalmente, en la publicación de los estudios de evaluación de un único marcador tumoral de interés, incluyendo a menudo su ajuste respecto a las variables pronósticas clínicas estándar. Estas directrices son especialmente relevantes para los estudios en los que se evalúa más de un marcador, pero no abordan las consideraciones estadísticas relativas al desarrollo de modelos pronósticos para un número elevado de marcadores candidato.

Publicación y otras cuestiones

Las recomendaciones REMARK se publicaron en septiembre de 2005 en cinco revistas de oncología: *British Journal of Cancer*, *European Journal of Cancer*, *Journal of Clinical Oncology*, *Journal of the National Cancer Institute*²⁵ y *Nature Clinical Practice Oncology*. Como ya se ha señalado, en ese momento todavía no se había completado el documento explicativo complementario en el que se recogen la significación y el fundamento de cada uno de los puntos con una revisión breve de la evidencia respecto a éstos; no obstante, se espera que este documento se publique lo antes posible. La lista de directrices REMARK se puede conocer a través de internet en la dirección <http://www.cancerdiagnosis.nci.nih.gov/assessment/progress/clinical.html>

Las directrices REMARK se desarrollaron específicamente para su aplicación en los estudios pronósticos sobre marcadores tumorales, aunque la mayor parte de las recomendaciones se pueden aplicar de la misma manera a los estudios pronósticos de carácter general. Ciertamente, aunque estos estudios son especialmente frecuentes en el ámbito de la oncología, también son muy habituales en todas las ramas de la medicina; en ambos casos, hay pruebas de la escasa calidad de los estudios pronósticos^{26,27}.

Comentarios

Los grupos STARD y REMARK se han fundamentado en los objetivos y procedimientos correspondientes al Grupo CONSORT. Sin embargo, ha habido algunas diferencias entre estos grupos, por ejemplo, en lo relativo a la falta de recursos o a la existencia de innovaciones de carácter genuino que podrían influir en la forma con que se lleve a cabo la próxima revisión CONSORT.

TABLA 2

Recomendaciones para la publicación de los estudios realizados sobre el valor pronóstico de los marcadores tumorales (REMARK, *REporting recommendations for tumour MARKer prognostic studies*)²⁵

	Punto n°	Descriptor	Informado en página número
Introducción	1	Indique el marcador evaluado, los objetivos del estudio y cualquier hipótesis pree especificada	
Métodos	2	Describa las características (p. ej., fase de la enfermedad, cuadros de comorbilidad) de los pacientes que participan en el estudio	
Participantes	3	Describa los tratamientos aplicados y las razones de su aplicación (p. ej., asignación aleatorizada o sistemática)	
Características de la muestra	4	Describa los materiales biológicos utilizados (incluyendo los controles), así como los métodos de conservación y almacenamiento	
	5	Especifique el tipo de ensayo utilizado y aporte (o cite) un protocolo detallado, incluyendo reactivos o dispositivos, procedimientos de control de calidad, evaluaciones de la reproducibilidad, los métodos cuantitativos y protocolos de evaluación y publicación de los resultados. Especifique si las determinaciones se efectuaron manteniendo enmascarada la variable respuesta	
Diseño del estudio	6	Especifique los métodos de selección de pacientes (prospectivo o retrospectivo) y los de estratificación o apareamiento empleados (p. ej., según la fase de la enfermedad, la edad, etc.). Especifique el período de selección de los casos, la fecha final y el tiempo de seguimiento medio	
	7	Defina con precisión todos los criterios de valoración clínica evaluados	
	8	Especifique todas las variables inicialmente evaluadas o consideradas para su inclusión en los modelos	
	9	Justifique el tamaño de la muestra. Si el estudio fue diseñado para detectar cierto tamaño del efecto, cuantifíquelo y describa la potencia deseada	
Análisis estadístico	10	Especifique los métodos estadísticos, incluyendo los criterios de selección de variables en la elaboración de modelos, los métodos de valoración de las premisas implícitas y el control de los datos ausentes	
	11	Especifique cómo han sido tratados los marcadores. Si es relevante, describa los métodos utilizados para determinar los valores umbral (<i>cutpoint</i>)	
Resultados	12	Describa el flujo de pacientes (puede ser útil un diagrama), incluyendo el número que participó en cada fase y las razones para el abandono del estudio. Mencione el número de pacientes evaluados y el número de episodios detectados tanto en conjunto como en cada subgrupo	
Datos	13	Describa las variables demográficas (al menos edad y sexo) y pronósticas de la enfermedad estudiada, así como el marcador tumoral. Incluya las cifras de datos ausentes	
Análisis y presentación	14	Muestre la relación entre el marcador y las variables pronósticas	
	15	Presente análisis univariados entre el marcador y la respuesta, mostrando la relación y el efecto estimado (p. ej., riesgo relativo y probabilidad de supervivencia). En lo posible, aporte análisis similares en todas las variables evaluadas. Para la relación entre el marcador tumoral y una variable de tipo "tiempo hasta...", se recomienda el uso de curvas de Kaplan-Meier	
	16	Sobre los análisis multivariados más importantes, presente los efectos estimados (p. ej., riesgo relativo) con sus intervalos de confianza entre el marcador y las restantes variables, al menos, las del modelo final	
	17	Con independencia de su significación, aporte los efectos estimados y sus intervalos de confianza en un análisis que considere todas las variables pronósticas estándar	
	18	Si procede, informe del análisis de las premisas, sensibilidad y validez interna	
Discusión	19	Interprete los resultados de acuerdo con las hipótesis pree especificadas y las resultantes de otros análisis exploratorios	
	20	Discuta el valor clínico de los resultados y las implicaciones para la investigación futura	

Es destacable el hecho de que las directrices de estos dos grupos, así como de CONSORT, QUOROM y otros, han necesitado años desde su inicio hasta su realización. Los intentos de mejora de la calidad de las publicaciones son importantes, pero a menudo son procesos lentos.

Como ya se ha señalado, las directrices de publicación como las señaladas en este artículo no representan algo fijo, sino que deben revisarse periódicamente a la vista de las nuevas evidencias. Esperamos que ambas directrices sean revisadas en los años venideros.

Agradecimientos

Las declaraciones STARD y REMARK no se podrían haber publicado sin el importante esfuerzo realizado por muchas personas, especialmente por las que aparecen en las listas de autores de las publicaciones.

REFERENCIAS BIBLIOGRÁFICAS

1. Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA*. 2001;285:1987-91.
2. Guyatt GH, Tugwell PX, Feeny DH, Haynes RB, Drummond M. A framework for clinical evaluation of diagnostic technologies. *CMAJ*. 1986;134: 587-94.
3. Knotnerus JA, Muris JW. Assessment of the accuracy of diagnostic tests: the cross-sectional study. *J Clin Epidemiol*. 2003;56:1118-28.
4. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med*. 2004;140:189-202.
5. Lijmer JG, Mol BW, Heisterkamp S, Bonsel 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.
6. 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.
7. Smidt N, Rutjes AW, Van der Windt DA, Ostelo RW, Reitsma JB, Bossuyt PM, et al. Quality of reporting of diagnostic accuracy studies. *Radiology*. 2005;235:347-53.

8. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Clin Chem.* 2003;49:1-6.
9. 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. *Clin Chem.* 2003;49:7-18.
10. Smidt N, Rutjes AWS, Van der Windt DAWM, Ostelo RWJG, Reitsma JB, Bossuyt PM, et al. Does the STARD Statement improve the quality of reporting of diagnostic accuracy studies? Fifth International Congress on Peer Review and Biomedical Publication. Chicago: September 16-18 2005.
11. Schilsky RL, Taube SE. Tumor markers as clinical cancer tests are we there yet? *Semin Oncol.* 2002;29:211-2.
12. McShane LM, Aamodt R, Cordon-Cardo C, Cote R, Faraggi D, Fradet Y, et al. Reproducibility of p53 immunohistochemistry in bladder tumors. National Cancer Institute, Bladder Tumor Marker Network. *Clin Cancer Res.* 2000;6:1854-64.
13. Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using «optimal» cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst.* 1994;86:829-35.
14. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer.* 1994;69:979-85.
15. Biganzoli E, Boracchi P, Marubini E. Biostatistics and tumor marker studies in breast cancer: design, analysis and interpretation issues. *Int J Biol Markers.* 2003;18:40-8.
16. Hall PA, Going JJ. Predicting the future: a critical appraisal of cancer prognosis studies. *Histopathology.* 1999;35:489-94.
17. Mirza AN, Mirza NQ, Vlastos G, Singletary SE. Prognostic factors in node-negative breast cancer: a review of studies with sample size more than 200 and follow-up more than 5 years. *Ann Surg.* 2002;235:10-26.
18. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Jones DR, Heney D, et al. Reporting of prognostic markers: current problems and development of guidelines for evidence-based practice in the future. *Br J Cancer.* 2003;88:1191-8.
19. Riley RD, Heney D, Jones DR, Sutton AJ, Lambert PC, Abrams KR, et al. A systematic review of molecular and biological tumor markers in neuroblastoma. *Clin Cancer Res.* 2004;10:4-12.
20. Burton A, Altman DG. Missing covariate data within cancer prognostic studies: a review of current reporting and proposed guidelines. *Br J Cancer.* 2004;91:4-8.
21. Popat S, Matakidou A, Houlston RS. Thymidylate synthase expression and prognosis in colorectal cancer: a systematic review and meta-analysis. *J Clin Oncol.* 2004;22:529-36.
22. Brundage MD, Davies D, Mackillop WJ. Prognostic factors in non-small cell lung cancer: a decade of progress. *Chest.* 2002;122:1037-57.
23. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Quality of Reporting of Meta-analyses.* *Lancet.* 1999;354:1896-900.
24. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med.* 2001;134:663-94.
25. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor marker prognostic studies (RE-MARK). *J Natl Cancer Inst.* 2005;97:1180-4.
26. Marx BE, Marx M. Prognosis of idiopathic membranous nephropathy: a methodologic meta-analysis. *Kidney International.* 1997;51:873-9.
27. Carson CA, Fine MJ, Smith MA, Weissfeld LA, Huber JT, Kapoor WN. Quality of published reports of the prognosis of community-acquired pneumonia. *J Gen Intern Med.* 1994;9:13-9.