

NOTA METODOLÓGICA

Instrumento para la lectura crítica y la evaluación de estudios epidemiológicos transversales

Silvina Berra^{a,b} / Josep Maria Elorza-Ricart^c / Maria-Dolors Estrada^{a,b} / Emília Sánchez^{a,b}

^aAgència d'Avaluació de Tecnologia i Recerca Mèdiques (AATRM), Barcelona, España;

^bCIBER Epidemiología y Salud Pública (CIBERESP), España; ^cEAP Gòtic, Institut Català de la Salut, Barcelona, España.

(Aroot for the critical appraisal of epidemiological cross-sectional studies)

Resumen

El objetivo de este trabajo fue desarrollar un instrumento para la evaluación de la calidad de los estudios epidemiológicos transversales. Se tuvieron en cuenta diferentes recomendaciones e instrumentos de valoración de estudios observacionales, como los de la Agency for Healthcare Research and Quality, el Scottish Intercollegiate Guidelines Group, el Osteba (Servicio de Evaluación de Tecnologías Sanitarias del País Vasco) y la iniciativa STROBE. El instrumento consta de 27 ítems para evaluar: pregunta u objetivo de investigación, participantes, comparabilidad entre los grupos estudiados, definición y medición de las variables principales, análisis y confusión, resultados, conclusiones, validez externa y aplicabilidad de los resultados, y conflicto de interés. Este instrumento puede utilizarse para la lectura crítica de artículos originales o en la elaboración de revisiones sistemáticas de la evidencia científica.

Palabras clave: Lectura crítica. Estudios transversales. Revisión sistemática.

Abstract

The aim was to develop a tool for the critical appraisal of epidemiological cross-sectional studies. Several recommendations or guidelines for assessing the strength of scientific evidence provided by observational studies were reviewed, like those from the Agency for Healthcare Research and Quality, the Scottish Intercollegiate Guidelines Group, the Osteba (Basque Office for Health Technology Assessment), and the STROBE Initiative. The tool has 27 items to assess: study question or objective, participants, comparability between groups, definition and measure of main variables; analysis and confusion, results, conclusions, external validity and applicability, and conflict of interest. This tool can be used to critically appraise research papers or to rate evidence during the elaboration of systematic reviews.

Key words: Critical appraisal. Cross-sectional studies. Systematic review.

Introducción

La evaluación y la síntesis del mejor conocimiento disponible sobre un tema forma parte actualmente de la práctica profesional habitual. Éste es un aspecto primordial del método empleado en la elaboración de una revisión sistemática de la evidencia científica (RSEC). El propósito de la lectura y la evaluación crítica de la literatura (*critical appraisal*) es clasificar la evidencia proporcionada por cada una de las publicaciones, según niveles jerárquicos, considerando tanto el diseño como el rigor metodológico del estudio.

Correspondencia: Silvina Berra.
Agència d'Avaluació de Tecnologia
i Recerca Mèdiques (AATRM).
Roc Boronat, 81-95, 2.^a planta. 08005 Barcelona. España.
Correo electrónico: sberra@aatrm.catsalut.net

Recibido: 28 de diciembre de 2007.

Aceptado: 4 de febrero de 2008.

Así, en investigaciones sobre la eficacia de una determinada intervención, un estudio clínico aleatorizado ofrece una evidencia superior a un estudio de casos y controles. De igual manera, un estudio clínico con baja probabilidad de sesgos debe proporcionar un mayor nivel de evidencia respecto a uno mal diseñado y que podría mostrar resultados no válidos.

Si tenemos en cuenta que la toma de decisiones se basa en los resultados de los diferentes estudios, o en las conclusiones de las RSEC o los metaanálisis que las incluyen, salta a la vista la importancia de disponer de instrumentos que evalúen y permitan clasificar de forma adecuada la evidencia. Por ello, se han ido desarrollando numerosos sistemas e instrumentos con el objetivo de evaluar la calidad de la evidencia¹. Los instrumentos (*checklists*) incluyen ítems o preguntas para guiar la lectura crítica y analizar la validez interna, los resultados y la validez externa de los estudios.

Por su génesis en la necesidad de contribuir a la toma de decisiones en la práctica clínica, estos instru-

mentos se han dirigido a estudios de carácter analítico, como los estudios clínicos aleatorizados, los estudios de cohortes y los estudios de casos y controles, así como también a estudios de pruebas diagnósticas, evaluaciones económicas y series de casos, además de a las mismas revisiones sistemáticas¹. En cambio, para otros tipos de diseños epidemiológicos, como los transversales, no se han desarrollado instrumentos de evaluación tan sistematizados.

Muchos problemas de interés en el campo de la salud pública, como conocer la frecuencia de un determinado estado de salud en la población, las desigualdades sociales en salud, el acceso a los servicios sanitarios o la evaluación de intervenciones preventivas, se abordan, a menudo, mediante estudios epidemiológicos descriptivos transversales. En este sentido, se han señalado las limitaciones de los estudios aleatorizados² y se ha indicado la necesidad de evaluar la credibilidad y el impacto de ciertas intervenciones mediante estudios con grupo de comparación y buen control de la confusión³. Por otra parte, algunos estudios transversales podrían tratarse como analíticos (cohortes o casos y controles) si plantean una asociación de determinación entre variables⁴. Por ejemplo, algunas variables, como la clase social o la migración, pueden considerarse determinantes de la salud o del acceso a los servicios sanitarios en una investigación de diseño transversal.

El objetivo de este trabajo fue desarrollar un instrumento para la evaluación de la calidad de estudios epidemiológicos transversales. El instrumento debe facilitar la lectura crítica y la valoración de los elementos respecto a la validez interna, la precisión y la utilidad de sus resultados.

Métodos

Se definen como epidemiológicos transversales los estudios en que las mediciones de las variables de interés se realizan en un mismo momento para las unidades de estudio incluidas^{5,6}. Habitualmente, estos estudios tienen la finalidad de estimar la frecuencia (prevalencia) de un estado o situación, o una medida de asociación (*odds ratio*, o razón de prevalencias) entre una exposición y un resultado, aunque no se disponga de seguimientos temporales⁶.

En la elaboración de este instrumento se tuvieron en cuenta las recomendaciones para la evaluación de la fuerza de la evidencia proveniente de estudios observacionales, según las conclusiones de la Agency for Healthcare Research and Quality¹ y los ítems incluidos en otros instrumentos de uso más frecuente en España, como los del Scottish Intercollegiate Guidelines Group⁷ y Osteba (Servicio de Evaluación de Tecnolo-

gías Sanitarias del País Vasco)⁸. Además, se tuvieron en consideración las recomendaciones para procesos editoriales de revisión de artículos epidemiológicos de la iniciativa STROBE^{6,9}. Los criterios considerados importantes se formularon como enunciados afirmativos.

Aplicación práctica y discusión

El instrumento tiene en total 27 ítems en las siguientes dimensiones de la evaluación de la evidencia científica (tabla 1): *a*) pregunta u objetivo de investigación (1 ítem); *b*) participantes (5 ítems); *c*) comparabilidad entre los grupos estudiados (4 ítems); *d*) definición y medición de las variables principales (4 ítems); *e*) análisis estadístico y confusión (4 ítems); *f*) resultados (4 ítems); *g*) conclusiones, validez externa y aplicabilidad de los resultados (4 ítems), y *h*) conflicto de intereses (1 ítem).

Sobre cada ítem el lector anota en qué medida se cumple tal criterio («muy bien», «bien», «regular» o «mal»), o si este aspecto «no está informado» o «no aplica» para el estudio en cuestión. Además, se ha confeccionado un enunciado resumen para cada dimensión y un enunciado para la evaluación global sobre la validez interna del estudio, que comprende las dimensiones «b» a «e». Finalmente, a partir de los enunciados resumen, el evaluador puede hacer una calificación global de la calidad del estudio. Para ello, se recomienda tener en cuenta las siguientes premisas: la calidad del estudio es «alta» si la mayoría de los enunciados resumen se responden como «muy bien» o «bien»; «media» si la validez interna es calificada como «regular» o la mayoría de los enunciados resumen se responden como «bien» o «regular»; y «baja» si la validez interna es calificada como «mal» o la mayoría de los enunciados resumen se responden como «regular» o «mal». Esta clasificación global puede utilizarse para decidir la inclusión o exclusión de artículos en una RSEC, quizás descartando los considerados de «baja» calidad.

En el futuro debe llevarse a cabo un análisis de la validez de criterio del instrumento, especialmente para recomendar su utilización y excluir los artículos de baja calidad, así como también su fiabilidad intra/interobservador. Por el momento, este desconocimiento es una limitación importante del presente trabajo.

Inicialmente este instrumento se desarrolló con la finalidad de cubrir la necesidad de evaluar artículos originales sobre la salud, el acceso y el uso de los servicios sanitarios de la población inmigrante respecto a la población autóctona, pero se concibió genéricamente, de manera que pueda ser útil para leer de forma crítica artículos originales de estudios epidemiológicos transversales sobre otros temas. Asimismo, se podría avan-

Tabla 1. Instrumento para la lectura crítica y la evaluación de estudios epidemiológicos transversales

	El aspecto se logra:				No informa	No aplica
	Muy bien	Bien	Regular	Mal		
a. Pregunta u objetivo de investigación						
1. En la formulación de la pregunta o del objetivo se menciona adecuadamente la población de estudio, las variables principales (independientes y dependientes) y el tipo de relación/comparación entre ellas						
<i>En resumen, el estudio se basa en una pregunta de investigación claramente definida</i>						
b. Participantes						
2. Se indican los criterios de inclusión y de exclusión de participantes, así como las fuentes y los métodos de selección						
3. Los criterios de elección son adecuados para dar respuesta a la pregunta o el objetivo del estudio						
4. La población de estudio, definida por los criterios de selección, contiene un espectro adecuado de la población de interés: Considerar en qué medida la población de estudio es representativa de toda la población de interés (población general, de escolares, etc.). Observar si grupos específicos dentro de esa población de estudio (p. ej., por nivel de instrucción o de formación, por ocupación, por país de procedencia, etc.) están proporcionalmente representados. Si el estudio se realiza en usuarios para luego inferir los resultados a una población mayor, este punto no está bien cubierto						
5. Se hizo una estimación del tamaño, el nivel de confianza o la potencia estadística de la muestra para la estimación de las medidas de frecuencia o de asociación que pretendía obtener el estudio						
6. Se informa del número de personas potencialmente elegibles, las inicialmente seleccionadas, las que aceptan y las que finalmente participan o responden. Si se comparan grupos, se indica esta información para cada grupo						
<i>En resumen, la muestra es adecuada y similar a la población base; se minimiza la posibilidad de sesgo de selección</i>						
c. Comparabilidad entre los grupos estudiados						
Si no se comparan grupos, responder «no aplica» a todos los enunciados de esta dimensión						
7. Las características de los grupos que se comparan están bien descritas. Por ejemplo, si se estudia un problema de salud, deben describirse los grupos por características sociodemográficas y otras variables que podrían modificar los resultados						
8. Las poblaciones de origen de los participantes de cada grupo son semejantes. Según la selección, ambas poblaciones tienen características similares, de tal manera que sean comparables en todo, excepto en el factor de estudio o de clasificación en uno u otro grupo						
9. Se utilizaron las mismas estrategias y técnicas de medición en todos los grupos; se midieron las mismas variables en todos los grupos						
10. No se produjeron pérdidas (por falta de medición, abandono, migración, etc.) que afecten a una parte de la muestra. Arbitrariamente, se podría considerar como alta una pérdida del 20% de la muestra; las pérdidas no deberían afectar al tamaño muestral mínimo necesario y sus causas no deberían ser diferentes entre los grupos						
<i>En resumen, los grupos estudiados son comparables; se minimiza la posibilidad de sesgo de selección</i>						

Tabla 1. Instrumento para la lectura crítica y la evaluación de estudios epidemiológicos transversales (continuación)

	El aspecto se logra:				No informa	No aplica
	Muy bien	Bien	Regular	Mal		
d. Definición y medición de las variables principales						
11. Se exponen claramente cuáles son las variables de exposición, resultado, confusoras o modificadoras						
12. Las variables principales tienen una adecuada definición conceptual (teórica) y operacional (escala de medición, sistema de clasificación, criterios diagnósticos, etc.)						
13. Los instrumentos de medición de las variables principales tienen validez y fiabilidad conocidas y adecuadas (se citan estudios que lo analizaron); se han adaptado culturalmente si las versiones originales provienen de lugares con lenguas o culturas diferentes (se citan los estudios que lo hicieron)						
14. Las técnicas de medición de las variables principales se describen suficientemente, son adecuadas y –si aplica– son las mismas para los grupos. Considerar la posibilidad de sesgos de memoria (alguno de los grupos puede recordar mejor algo del pasado) o del entrevistador (por conocimiento de la exposición o del problema de salud)						
<i>En resumen, la medición de las variables principales se realizó de forma adecuada; se minimiza la posibilidad de sesgos de información</i>						
e. Análisis estadístico y confusión*						
15. El análisis estadístico estuvo determinado desde el inicio del estudio						
16. Se especifican las pruebas estadísticas utilizadas y son adecuadas						
17. Se trataron correctamente las pérdidas de participantes, datos perdidos u otros efectos del diseño de la muestra (diferentes probabilidades de selección) o de la exclusión de casos para algunos análisis						
18. Se tuvieron en cuenta los principales elementos de confusión posibles en el diseño y en el análisis En el diseño deberían incorporarse variables teóricamente asociadas o determinantes del problema estudiado. En el análisis, la estimación del resultado principal debería estratificarse o ajustarse por esas variables						
<i>En resumen, el análisis es adecuado y se minimiza la posibilidad de confusión</i>						
Valoración global de la validez interna Considerar las dimensiones b-e	Muy bien	Bien	Regular	Mal		
<i>En resumen, el diseño del estudio permite minimizar los sesgos y el efecto de confusión</i>						
f. Resultados						
19. Se incluyen resultados de todos los participantes o se indica el número de datos no disponibles						
20. Se presentan los resultados planteados en los objetivos y todos los de interés, de manera clara y comprensible						
21. Se presentan medidas brutas y ajustadas, indicando las variables por las que se ajustan los resultados y justificando cuáles se incluyeron (o no) en el análisis						
22. Se presentan estimaciones de la significación estadística de las diferencias entre grupos (p. ej., valores de p) o de la precisión de los resultados (p. ej., intervalos de confianza)						
<i>En resumen, los resultados están bien descritos, son útiles y precisos</i>						

(Continúa)

Tabla 1. Instrumento para la lectura crítica y la evaluación de estudios epidemiológicos transversales (continuación)

	El aspecto se logra:				No informa	No aplica
	Muy bien	Bien	Regular	Mal		
g. Conclusiones, validez externa y aplicabilidad de los resultados						
23. Las conclusiones dan respuesta a los objetivos del estudio						
24. Las conclusiones presentadas se basan en los resultados obtenidos						
25. Los resultados de este estudio pueden extrapolarse a la población de interés de la presente revisión. Analizar similitudes y diferencias de ambas poblaciones (la del estudio y la de interés del lector) considerando el contexto espacial y temporal (p. ej., la prevalencia de la exposición), los criterios de inclusión, la definición y la medición de la exposición y el resultado, el nivel de confianza de las estimaciones, etc.						
26. La discusión considera implicaciones de la aplicación de los resultados, beneficios, seguridad y costes de su aplicación						
<i>En resumen, los resultados del estudio son generalizables a la población y contexto en que interesa aplicarlos</i>						
h. Conflicto de intereses						
27. Se menciona la fuente de financiación del estudio o los autores declaran la existencia o ausencia de conflictos de intereses						
<i>En resumen, los conflictos de intereses no condicionan los resultados ni las conclusiones del estudio</i>						
Valoración global de la calidad del estudio	Alta	Media	Baja			
La calidad de la evidencia aportada por el estudio es ^b						

^aSi bien la definición de confusión implica una relación causal, se utiliza este término para indicar la necesidad de tener en cuenta otras variables que pueden modificar el estimador de la asociación estudiada.

^bComo orientación, la calidad del estudio se puede considerar alta si la mayoría de los enunciados resumen se responden como «muy bien» o «bien»; media si la validez interna es calificada como «regular», o la mayoría de los enunciados resumen se responden como «bien» o «regular», y baja si la validez interna es calificada como «mal», o la mayoría de los enunciados resumen se responden como «regular» o «mal».

zar, mediante algunas adaptaciones, en el diseño de un instrumento similar para estudios descriptivos longitudinales (p. ej., series de casos) o estudios casi experimentales (p. ej., estudios antes-después o ensayos comunitarios no aleatorizados).

Agradecimientos

Este trabajo se ha realizado en el marco de colaboración previsto en el Plan de Calidad para el Sistema Nacional de Salud, al amparo del convenio de colaboración suscrito por el Instituto de Salud Carlos III, organismo dependiente del Ministerio de Sanidad y Consumo, y la Agència d'Avaluació de Tecnologia i Recerca Mèdiques (AATRM). Proyecto: «Salud y uso de los servicios sanitarios de inmigrantes en España. Revisión sistemática de la evidencia científica».

Bibliografía

- West S, King V, Carey TS, et al. Systems to rate the strength of scientific evidence. Evidence report/technology assessment N.º 47 (prepared by the Research Triangle Institute-University of North Carolina Evidence-based Practice Center under Contract N.º 290-97-0011). AHRQ Publication N.º 02-E016. Rockville: Agency for Healthcare Research and Quality; 2002.
- Victora CG, Habicht JP, Bryce J. Evidence-based public health: moving beyond randomized trials. Am J Public Health. 2004;94:400-5.
- Habicht JP, Victora CG, Vaughan JP. Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. Int J Epidemiol. 1999;28:10-8.
- Vandenbroucke JP, Von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al, for the STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med. 2007;4:e297.
- Last JM, editor. A dictionary of epidemiology. 4th ed. New York: Oxford University Press; 2001.

6. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, en nombre de la Iniciativa STROBE. Declaración de la Iniciativa STROBE (Strengthening the Reporting of Observational Studies in Epidemiology): directrices para la comunicación de estudios observacionales. *Gac Sanit.* 2008;22:144-50.
7. Scottish Intercollegiate Guidelines Group. A guideline developers' handbook. Edinburgh: SIGN; 2001. Publication N.º 50.
8. López de Argumedo M, Rico R, Andrio E, Reviriego E, Hurtado de Saracho I, Asua J. *OstFLCrítica*. Fichas de lectura crítica de la literatura científica. Vitoria-Gasteiz: Osteba-Servicio de Evaluación de Tecnologías Sanitarias. Departamento de Sanidad. Gobierno Vasco; 2006.
9. Fernández E, Cobo E, Guablar-Castillón P. La declaración STROBE o cómo mejorar la presentación de los estudios observacionales. *Gac Sanit.* 2008;22:87-9.