

Artículo especial

e Mejorar la comunicación de estudios observacionales en epidemiología (STROBE): explicación y elaboración

Jan P. Vandenbroucke^a, Erik Von Elm^{b,c,*}, Douglas G. Altman^d, Peter C. Gøtzsche^e, Cynthia D. Mulrow^f, Stuart J. Pocock^g, Charles Poole^h, James J. Schlesselmanⁱ, Matthias Egger^{bj} y en nombre de la Iniciativa STROBE

^a Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

^b Institute of Social & Preventive Medicine (ISPM), University of Bern, Bern, Switzerland

^c Department of Medical Biometry and Medical Informatics, University Medical Centre, Freiburg, Germany

^d Cancer Research UK/NHS Centre for Statistics in Medicine, Oxford, United Kingdom

^e Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark

^f University of Texas Health Science Center, San Antonio, United States of America

^g Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London, United Kingdom

^h Department of Epidemiology, University of North Carolina School of Public Health, Chapel Hill, United States of America

ⁱ Department of Biostatistics, University of Pittsburgh Graduate School of Public Health, and University of Pittsburgh Cancer Institute, Pittsburgh, United States of America

^j Department of Social Medicine, University of Bristol, Bristol, United Kingdom

INFORMACIÓN DEL ARTÍCULO

On-line el 26 de febrero de 2009

RESUMEN

Gran parte de la investigación biomédica es de tipo observacional. Los informes de los estudios observacionales a menudo poseen una calidad insuficiente, lo que dificulta la evaluación de sus fortalezas y debilidades para generalizar los resultados. Teniendo en cuenta la evidencia empírica y consideraciones teóricas, un grupo de expertos en metodología, investigadores y editores de revistas científicas, desarrollaron una lista de recomendaciones para aumentar la calidad de las publicaciones de los estudios observacionales: *Strengthening the Reporting of Observational Studies in Epidemiology* (STROBE). La Declaración STROBE consiste en una lista de verificación de 22 puntos que guardan relación con las diferentes secciones de un artículo: título, resumen, introducción, metodología, resultados y discusión. De ellos, 18 puntos son comunes a los tres diseños de estudio: cohorte, casos y controles, y transversales; los otros cuatro son específicos para cada una de estas tres modalidades. La Declaración STROBE proporciona a los autores información sobre cómo mejorar la calidad de los artículos sobre estudios observacionales y facilita a los revisores, editores de revistas y lectores su apreciación crítica y su interpretación. Este documento explicativo tiene el propósito de impulsar el uso, la comprensión y la difusión de la Declaración STROBE. Se presentan el significado y el análisis razonado para cada punto de la lista de verificación, proporcionando uno o varios ejemplos publicados en la literatura y, en lo posible, referencias de estudios empíricos relevantes y literatura metodológica. También se incluyen ejemplos de diagramas de flujo. La Declaración STROBE, el presente documento y la página Web asociada (<http://www.strobe-statement.org/>) son recursos útiles para mejorar la divulgación de la investigación observacional.

© 2007 Publicado por Elsevier España, S.L.

Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration

ABSTRACT

Much medical research is observational. The reporting of observational studies is often of insufficient quality. Poor reporting hampers the assessment of the strengths and weaknesses of a study and the generalisability of its results. Taking into account empirical evidence and theoretical considerations, a group of methodologists, researchers, and editors developed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) recommendations to improve the quality of reporting of observational studies. The STROBE Statement consists of a checklist of 22 items, which relate to the title, abstract, introduction, methods, results and discussion sections of articles. Eighteen items are common to cohort studies, case-control studies and cross-sectional studies and four are specific to each of the three study designs. The STROBE Statement provides guidance to authors about how to improve the reporting of observational studies and facilitates critical appraisal and interpretation of studies by reviewers, journal editors and readers. This explanatory and elaboration document is intended to enhance the use, understanding, and dissemination of the STROBE Statement. The meaning and rationale for each checklist item are presented. For each item, one or several published examples and, where possible, references to relevant empirical studies and methodological literature are provided. Examples of useful flow diagrams are also included. The STROBE Statement, this document, and the associated Web site (<http://www.strobe-statement.org/>) should be helpful resources to improve reporting of observational research.

© 2007 Published by Elsevier España, S.L.

Introducción

La práctica racional del cuidado de la salud requiere del conocimiento sobre la etiología, la patogénesis, el diagnóstico, el

* Autor para correspondencia.

Correo electrónico: strobe@ispm.unibe.ch (E. Von Elm).

pronóstico y el tratamiento de las enfermedades. Los ensayos clínicos aleatorizados proporcionan una valiosa evidencia sobre los tratamientos y otras intervenciones. Sin embargo, gran parte del conocimiento clínico o de salud pública proviene de la investigación observacional¹. Nueve de cada 10 artículos de investigación publicados en revistas de especialidades clínicas describen investigaciones observacionales^{2,3}.

La declaración STROBE

Con frecuencia, la comunicación de estudios observacionales no es lo suficientemente clara y detallada para evaluar las fortalezas y las debilidades de la investigación^{4,5}. Para mejorarla, desarrollamos una lista de verificación de los puntos que se deben atender: la Declaración STROBE, que recientemente se ha publicado en varias revistas científicas⁶. Estos puntos se relacionan con las secciones de título, resumen, introducción, métodos, resultados y discusión. Nuestro propósito es asegurar una presentación clara de lo que se planeó, realizó y encontró en un estudio observacional. Hacemos énfasis en que estas recomendaciones no son recetas para planificar o conducir estudios y tampoco dictan la metodología ni decretan una presentación uniforme.

La declaración STROBE proporciona recomendaciones generales para estudios observacionales descriptivos y aquellos que investigan asociaciones entre variables de exposición y resultados de salud. Comprende los tres tipos principales de estudios observacionales: de cohortes, de casos y controles, y transversales. Los autores utilizan terminología diversa para describir estos diseños. Por ejemplo, “estudio de seguimiento” y “estudio longitudinal” se utilizan como sinónimos para el estudio de cohortes, y “estudio de prevalencia” como sinónimo de “estudio transversal”. Elegimos la actual terminología porque es de uso

Recuadro 1—Principales diseños de estudio tratados por STROBE.

Los diseños de cohortes, casos y controles, y transversales, constituyen diferentes enfoques para investigar la ocurrencia de eventos relacionados con la salud en una población y en un periodo de tiempo determinados. Estos estudios pueden abordar muchos tipos de eventos relacionados con la salud, incluyendo la enfermedad o su remisión, la discapacidad o las complicaciones, la muerte o la supervivencia, así como la presencia de factores de riesgo.

En los estudios de cohortes, los investigadores siguen a las personas durante un periodo de tiempo. Obtienen información acerca de ellas y de su exposición basal, dejan que pase el tiempo y posteriormente comprueban la presencia de eventos resultado. Los investigadores comúnmente hacen comparaciones entre los individuos expuestos y no expuestos, o bien entre grupos de individuos con diferentes categorías de exposición. También pueden medir varios eventos diferentes y examinar variables de exposición y eventos en diversos momentos durante el periodo de seguimiento. Las cohortes cerradas (por ejemplo, cohortes desde el nacimiento) reclutan un número definido de participantes al inicio del estudio y los siguen de ahí en adelante, a menudo a intervalos predeterminados, hasta una fecha final también fijada. En las cohortes abiertas, la población en estudio es dinámica: las personas entran y salen del estudio en diferentes momentos en el tiempo (por ejemplo, los habitantes de una comunidad). Las cohortes abiertas cambian debido a muertes, nacimientos y migración, pero la composición de la población respecto a variables

como edad y sexo puede permanecer aproximadamente constante, en especial en un periodo de tiempo corto. En una cohorte cerrada se pueden estimar las incidencias acumuladas (riesgos) y las tasas de incidencia; cuando se comparan los grupos de expuestos y no expuestos, es posible estimar la razón de riesgos o la razón de tasas. Las cohortes abiertas estiman tasas de incidencia y razones de tasas.

En los estudios de casos y controles, los investigadores comparan exposiciones entre personas con una enfermedad particular (casos) y personas sin esa enfermedad (controles). Reúnen casos y controles que sean representativos de una cohorte subyacente o de una muestra de una población. Esa población puede estar definida geográficamente, o provenir de la zona de influencia del centro de atención de salud. La muestra de casos puede ser del 100% o una fracción grande de casos disponibles, mientras que la muestra de controles generalmente es sólo una pequeña fracción de personas que no tienen la enfermedad en estudio. Los controles representan la cohorte o población de personas de la cual surgieron los casos. Los investigadores calculan la *odds ratio* de las exposiciones para las supuestas causas de la enfermedad entre los casos y los controles (véase el [Recuadro 7](#)). Dependiendo de la estrategia de muestreo para casos y controles, y de la naturaleza de la población estudiada, la *odds ratio* obtenida en un estudio de casos y controles se interpreta como la razón de riesgos, razón de tasas u *odds ratio* para la prevalencia^{16,17}. La mayoría de los estudios de casos y controles publicados muestrean cohortes abiertas, lo que les permite estimar directamente las razones de tasas.

En los estudios transversales, los investigadores estudian a todos los individuos de una muestra en el mismo momento en el tiempo, y con frecuencia investigan la prevalencia de exposiciones, factores de riesgo o enfermedades. Algunos estudios transversales son analíticos y están dirigidos a cuantificar posibles asociaciones causales entre exposiciones y enfermedades. Tales estudios pueden analizarse como un estudio de cohortes mediante la comparación de la prevalencia de enfermedad entre los grupos de exposición. También pueden analizarse como un estudio de casos y controles al comparar la razón de exposición entre los grupos con y sin enfermedad. Un problema que puede surgir en cualquier diseño epidemiológico, pero que es particularmente evidente en los estudios transversales, es establecer que la exposición precedió a la enfermedad aun cuando, algunas veces, el orden en el tiempo de la exposición y el evento pueda estar claro. Por ejemplo, en un estudio en que la variable de exposición es congénita o genética, podemos estar seguros de que la exposición precedió a la enfermedad, aunque estemos midiendo ambas al mismo tiempo.

común. Desafortunadamente, con frecuencia estos términos se utilizan de forma incorrecta⁷ o imprecisa⁸. En el [Recuadro 1](#) describimos las características distintivas de los tres diseños de estudio.

El ámbito de la investigación observacional

Los estudios observacionales sirven a una amplia gama de propósitos: desde la información sobre un primer indicio de la posible causa de una enfermedad, hasta la verificación de la magnitud de asociaciones previamente comunicadas. Las ideas para los estudios pueden surgir de observaciones clínicas o del refinamiento del conocimiento biológico, así como de observaciones informales de datos que conducen a mayores exploraciones. Del mismo modo que un clínico ve miles de pacientes y alguno llama su atención, el investigador puede fijarse en algo

especial en los datos observados. No es posible ni deseable corregir las interpretaciones mediante la revisión múltiple de los datos⁹, pero a menudo son necesarios nuevos estudios para confirmar o refutar las observaciones iniciales¹⁰.

Los datos existentes se pueden utilizar para investigar nuevas ideas sobre posibles factores causales, y pueden ser suficientes para rechazar o confirmar una hipótesis. En otros casos, posteriormente se diseñan estudios específicos para superar los problemas que pueden haberse encontrado en comunicaciones previas. Estos estudios recopilarán nuevos datos y se planificarán para ese propósito, en contraste con los análisis de datos existentes. Esto conduce a diversos puntos de vista, por ejemplo respecto a la formación de subgrupos o a la importancia de un tamaño de muestra predeterminado. STROBE intenta incluir estos diferentes usos de la investigación observacional, desde el descubrimiento hasta la refutación o la confirmación. Cuando sea necesario indicaremos en qué circunstancias específicas se aplican las recomendaciones.

Cómo utilizar este documento

Este documento está ligado a la versión breve de la Declaración STROBE, que presentó en varias revistas⁶ la lista de puntos a verificar y forma parte integral de la Declaración STROBE. Nuestra intención es explicar cómo informar adecuadamente de la investigación, no cómo ésta debe llevarse a cabo. Ofrecemos una explicación detallada para cada punto de la lista. Cada explicación va precedida por un ejemplo de lo que consideramos un informe adecuado. Esto no significa que el estudio de donde se ha tomado el ejemplo esté bien escrito en su totalidad, y tampoco que sus resultados sean fiables, ni que posteriormente sean confirmados por otros estudios; significa solamente que este punto en particular está bien escrito y comunicado. Además de explicaciones y ejemplos, incluimos los recuadros 1 a 8 con información adicional. Estos recuadros están dedicados a los lectores que deseen refrescar su memoria sobre algunos aspectos teóricos o quieran informarse rápidamente sobre detalles de cuestiones técnicas. Su completa comprensión puede requerir estudiar los libros de texto o los documentos metodológicos que se citan.

Las recomendaciones de STROBE no tratan específicamente temas como estudios de asociaciones genéticas, modelos de enfermedades infecciosas, informes o series de casos^{11,12}. Sin embargo, muchos de los elementos de STROBE pueden aplicarse a estos diseños, por lo que los autores que comuniquen tales estudios pueden encontrar útiles estas recomendaciones. Para los autores de artículos sobre estudios observacionales que traten específicamente pruebas diagnósticas, marcadores tumorales y asociaciones genéticas¹⁵, las recomendaciones STARD¹³, REMARK¹⁴ y STREGA pueden ser particularmente útiles.

Los puntos de la lista de verificación

A continuación comentamos y explicamos los 22 puntos de la lista de verificación STROBE (tabla 1), y ofrecemos ejemplos publicados para cada uno de ellos. Algunos de estos ejemplos se han reproducido eliminando referencias o abreviaturas. Dieciocho de los puntos se aplican a los tres diseños de estudio, mientras que los otros cuatro son específicos. Los puntos con asterisco (por ejemplo el número 8*) indican que la información se debe dar por separado para los casos y los controles en los estudios de casos y controles, y para los grupos expuestos y no expuestos en los estudios de cohortes y transversales. Aconsejamos a los autores tratar todos los puntos en alguna parte de su artículo, pero no indicamos una localización ni un orden exactos. Por ejemplo,

hablamos de la comunicación de resultados en puntos separados, pero reconocemos que los autores pueden tratar varios puntos dentro de una sola sección del texto o en una tabla.

Los puntos

Título y resumen

1 (a). Indique el diseño del estudio en el título o en el resumen con un término habitual

Ejemplo

“Incidencia de leucemia entre trabajadores de la industria manufacturera de zapatos y botas: un estudio de casos y controles”¹⁸.

Explicación

Los lectores deben poder identificar fácilmente, desde el título o el resumen, el diseño que se utilizó. Un término explícito utilizado habitualmente para el diseño del estudio, ayuda también a asegurar la correcta indexación del artículo en las bases de datos electrónicas^{19,20}.

1 (b). Proporcione en el resumen una sinopsis informativa y equilibrada de lo que se ha hecho y lo que se ha encontrado

Ejemplo

“Antecedentes: la supervivencia esperada en los pacientes infectados por el VIH es de fundamental interés para la Salud Pública.

Objetivo: estimar el tiempo de supervivencia y las tasas de mortalidad específicas para la edad de una población infectada por VIH en comparación con la mortalidad en la población general.

Diseño: estudio de cohortes de base poblacional.

Contexto: todas las personas infectadas con VIH que recibieron atención en Dinamarca de 1995 a 2005.

Pacientes: cada miembro del estudio de la cohorte nacional danesa de VIH se emparejó hasta con 99 personas de la población general según sexo, fecha de nacimiento y lugar de residencia.

Mediciones: los autores calcularon las tablas de vida de Kaplan-Meier con la edad como escala de tiempo para estimar la supervivencia desde los 25 años. Se observó a los pacientes infectados con VIH y a las personas correspondientes de la población general a partir de la fecha del diagnóstico del VIH hasta su muerte, migración o el 1 de mayo de 2005.

Resultados: se incluyeron en el estudio 3.990 pacientes infectados por el VIH y 379.872 personas de la población general, que produjeron 22.744 personas-año (mediana 5,8) y 2.689.287 personas-año (mediana 8,4) de observación. Se perdió el 3% de los participantes durante el seguimiento. A partir de los 25 años de edad, la mediana del tiempo de supervivencia para los pacientes infectados fue 19,9 años (IC95%: 18,5–21,3), y 51,1 años (IC95%: 50,9–51,5) para la población general. Para los pacientes infectados, la supervivencia aumentó a 32,5 años (IC95%: 29,4–34,7) durante el periodo 2000–2005. En el subgrupo que excluyó a personas con coinfección por el virus de la hepatitis C (16%), la mediana de la supervivencia fue 38,9 años (IC95%: 35,4–40,1) durante este mismo periodo. La tasa de mortalidad relativa para los pacientes con infección por VIH comparada con la de la población general disminuyó con el aumento de la edad, mientras que el exceso de la tasa de mortalidad se incrementó con la edad.

Tabla 1
Declaración STROBE: lista de puntos esenciales que deben describirse en la publicación de estudios observacionales

Título y resumen	Punto	Recomendación
Introducción	1	(a) Indique, en el título o en el resumen, el diseño del estudio con un término habitual (b) Proporcione en el resumen una sinopsis informativa y equilibrada de lo que se ha hecho y lo que se ha encontrado
Contexto/ fundamentos	2	Explique las razones y el fundamento científicos de la investigación que se comunica
Objetivos	3	Indique los objetivos específicos, incluyendo cualquier hipótesis preespecificada
Métodos		
Diseño del estudio	4	Presente al principio del documento los elementos clave del diseño del estudio
Contexto	5	Describa el marco, los lugares y las fechas relevantes, incluyendo los periodos de reclutamiento, exposición, seguimiento y recogida de datos
Participantes	6	(a) Estudios de cohortes: proporcione los criterios de elegibilidad, así como las fuentes y el método de selección de los participantes. Especifique los métodos de seguimiento Estudios de casos y controles: proporcione los criterios de elegibilidad, así como las fuentes y el proceso diagnóstico de los casos y el de selección de los controles. Indique las razones para la elección de casos y controles Estudios transversales: proporcione los criterios de elegibilidad, y las fuentes y los métodos de selección de los participantes (b) Estudios de cohortes: en los estudios pareados, proporcione los criterios para la formación de parejas y el número de participantes con y sin exposición Estudios de casos y controles: en los estudios pareados, proporcione los criterios para la formación de las parejas y el número de controles por cada caso
Variables	7	Defina claramente todas las variables: de respuesta, exposiciones, predictoras, confusoras y modificadoras del efecto. Si procede, proporcione los criterios diagnósticos
Fuentes de datos/medidas	8*	Para cada variable de interés, indique las fuentes de datos y los detalles de los métodos de valoración (medida). Si hubiera más de un grupo, especifique la comparabilidad de los procesos de medida
Sesgos	9	Especifique todas las medidas adoptadas para afrontar posibles fuentes de sesgo
Tamaño muestral	10	Explique cómo se determinó el tamaño muestral
Variables cuantitativas	11	Explique cómo se trataron las variables cuantitativas en el análisis. Si procede, explique qué grupos se definieron y por qué
Métodos estadísticos	12	(a) Especifique todos los métodos estadísticos, incluidos los empleados para controlar los factores de confusión (b) Especifique todos los métodos utilizados para analizar subgrupos e interacciones (c) Explique el tratamiento de los datos ausentes (<i>missing data</i>) (d) Estudios de cohortes: si procede, explique cómo se afrontan las pérdidas en el seguimiento Estudios de casos y controles: si procede, explique cómo se parearon casos y controles Estudios transversales: si procede, especifique cómo se tiene en cuenta en el análisis la estrategia de muestreo (e) Describa los análisis de sensibilidad
Resultados		
Participantes	13*	(a) Indique el número de participantes en cada fase del estudio; p. ej., número de participantes elegibles, analizados para ser incluidos, confirmados elegibles, incluidos en el estudio, los que tuvieron un seguimiento completo y los analizados (b) Describa las razones de la pérdida de participantes en cada fase (c) Considere el uso de un diagrama de flujo
Datos descriptivos	14*	(a) Describa las características de los participantes en el estudio (p. ej., demográficas, clínicas, sociales) y la información sobre las exposiciones y los posibles factores de confusión (b) Indique el número de participantes con datos ausentes en cada variable de interés (c) Estudios de cohortes: resuma el periodo de seguimiento (p. ej., promedio y total)
Datos de las variables de resultado	15*	Estudios de cohortes: indique el número de eventos resultado o bien proporcione medidas resumen a lo largo del tiempo Estudios de casos y controles: indique el número de participantes en cada categoría de exposición o bien proporcione medidas resumen de exposición Estudios transversales: indique el número de eventos resultado o bien proporcione medidas resumen
Resultados principales	16	(a) Proporcione estimaciones no ajustadas y, si procede, ajustadas por factores de confusión, así como su precisión (p. ej., intervalos de confianza del 95%). Especifique los factores de confusión por los que se ajusta y las razones para incluirlos (b) Si categoriza variables continuas, describa los límites de los intervalos (c) Si fuera pertinente, valore acompañar las estimaciones del riesgo relativo con estimaciones del riesgo absoluto para un periodo de tiempo relevante
Otros análisis	17	Describa otros análisis efectuados (de subgrupos, interacciones o sensibilidad)
Discusión		
Resultados clave	18	Resuma los resultados principales de los objetivos del estudio
Limitaciones	19	Discuta las limitaciones del estudio, teniendo en cuenta posibles fuentes de sesgo o de imprecisión. Razone tanto sobre la dirección como sobre la magnitud de cualquier posible sesgo
Interpretación	20	Proporcione una interpretación global prudente de los resultados considerando objetivos, limitaciones, multiplicidad de análisis, resultados de estudios similares y otras pruebas empíricas relevantes
Generabilidad	21	Discuta la posibilidad de generalizar los resultados (validez externa)
Otra información		
Financiación	22	Especifique la financiación y el papel de los patrocinadores del estudio, y si procede, del estudio previo en que se basa su artículo

Nota: Se ha publicado un artículo que explica y detalla la elaboración de cada punto de la lista, y se ofrece el contexto metodológico y ejemplos reales de comunicación transparente. La lista de puntos STROBE se debe utilizar preferiblemente junto con ese artículo (gratuito en las páginas web de las revistas *PLoS Medicine* (<http://www.plosmedicine.org/>), *Annals of Internal Medicine* (<http://www.annals.org/>) y *Epidemiology* (<http://www.epidem.com/>)). En la página web de STROBE (<http://www.strobe-statement.org>) aparecen las diferentes versiones de la lista correspondientes a los estudios de cohortes, a los estudios de casos y controles, y a los estudios transversales.

* Proporcione esta información por separado para casos y controles en los estudios con diseño de casos y controles Si procede, también para los grupos con y sin exposición en los estudios de cohortes y en los transversales.

Limitaciones: se asumió que las tasas de mortalidad observadas serían aplicables más allá del tiempo máximo de observación de 10 años.

Conclusiones: la mediana estimada del tiempo de supervivencia es de más de 35 años para una persona joven diagnosticada de infección por VIH en la última etapa del tratamiento anti-retroviral de gran actividad. Sin embargo, se necesita un esfuerzo continuo para reducir más las tasas de mortalidad de estas personas comparadas con la población general²¹.

Explicación

El resumen proporciona la información clave que permite al lector entender el estudio y decidir si lee o no el artículo. Los componentes típicos incluyen la pregunta de investigación, una breve descripción de los métodos y los resultados, y una conclusión²². Los resúmenes deben sintetizar detalles clave del estudio y presentar solamente la información que se proporciona en el artículo. Aconsejamos la presentación de los resultados principales en cifras, incluyendo el número de participantes, las estimaciones de asociaciones y las medidas apropiadas de la variabilidad y de la incertidumbre (por ejemplo *odds ratio* con intervalo de confianza). Creemos insuficiente indicar solamente si una exposición se asocia o no de manera significativa con el resultado.

Una serie de encabezados referentes a los antecedentes, el diseño, el desarrollo y el análisis del estudio pueden ayudar a los lectores a obtener rápidamente la información esencial²³. Muchas revistas requieren estos resúmenes estructurados, que tienden a ser de mayor calidad y más informativos que los no estructurados^{24,25}.

Introducción

La sección de introducción debe describir por qué se llevó a cabo el estudio y qué preguntas de investigación e hipótesis trata. Debe permitir que otros comprendan el contexto del estudio y que juzguen su posible contribución al conocimiento actual.

2. Contexto/fundamentos: explique las razones y el fundamento científicos de la investigación que se comunica

Ejemplo

“La preocupación por el aumento de la prevalencia de la obesidad en niños y adolescentes se ha centrado en las bien documentadas asociaciones entre la obesidad infantil y el aumento del riesgo cardiovascular y la mortalidad en la edad adulta. La obesidad infantil tiene considerables consecuencias sociales y psicológicas, tanto en la niñez como en la adolescencia; no obstante, poco se sabe sobre sus consecuencias sociales, socioeconómicas y psicológicas en la edad adulta. Una revisión sistemática reciente no encontró estudios longitudinales sobre las consecuencias de la obesidad infantil, a excepción de los efectos físicos sobre la salud, y solamente dos estudios longitudinales sobre los efectos socioeconómicos de la obesidad en la adolescencia. Gortmaker et al. encontraron que mujeres norteamericanas que habían sido obesas al final de la adolescencia, en 1981, tenían menos probabilidades de estar casadas y menores ingresos siete años después que las mujeres que no habían tenido sobrepeso, mientras que los hombres que tenían sobrepeso era menos probable que estuvieran casados. En Gran Bretaña, Sargent et al. encontraron que las mujeres, pero no los hombres, que habían sido obesas a los 16 años en 1974, ganaban un 7,4% menos que sus pares no obesas a la edad de 23 años. (...) Utilizamos datos longitudinales de la cohorte de nacimientos británicos de

1970 para examinar las consecuencias socioeconómicas, educativas, sociales y psicológicas, en la edad adulta, de la obesidad infantil²⁶.

Explicación

Los antecedentes científicos del estudio proporcionan un contexto importante a los lectores. Establecen la etapa en que se ubica el estudio y describen su propósito. Ofrecen una descripción de lo que se sabe acerca del tema y qué vacíos del conocimiento actual se tratan en el estudio. La información de los antecedentes debe centrarse en estudios recientes y en todas las revisiones sistemáticas de estudios pertinentes.

3. Objetivos: indique los objetivos específicos, incluyendo cualquier hipótesis preespecificada

Ejemplo

Nuestros objetivos principales fueron: 1) determinar la prevalencia de la violencia doméstica entre mujeres que solicitaron consulta médica en cuatro centros comunitarios de atención primaria que atienden a pacientes de diversos estratos socioeconómicos, y 2) identificar diferencias demográficas y clínicas entre las pacientes que sufrieron abusos recientemente y las que no²⁷.

Explicación

Los objetivos son los propósitos detallados del estudio. Los objetivos bien elaborados especifican la población, las exposiciones y los efectos, así como los parámetros que serán estimados. Pueden formularse como hipótesis específicas o como preguntas que el estudio pretende abordar. En algunas situaciones los objetivos pueden ser menos específicos, por ejemplo en las etapas tempranas de una investigación. Independientemente de ello, el informe debe reflejar claramente las intenciones de los investigadores. Por ejemplo, si durante el análisis de datos se distinguieron subgrupos importantes o se realizaron análisis no previstos con referencia al propósito original del estudio, es necesario incluir la descripción correspondiente (véase también los puntos 4, 17 y 20).

Métodos

La sección de métodos debe describir lo que se planeó y lo que se hizo con suficiente detalle para permitir a otros entender los aspectos esenciales del estudio, para juzgar si los métodos fueron adecuados para proporcionar respuestas fiables y válidas, así como para evaluar si cualquier desviación del plan original fue razonable.

4. Diseño del estudio: presente al principio del documento los elementos clave del diseño del estudio

Ejemplo

“Usamos un diseño de casos alternantes o cruzados (*case-crossover*), que es una variante del diseño de casos y controles útil cuando una exposición breve (uso del teléfono mientras se conduce) causa un incremento pasajero en el riesgo de un desenlace de poca frecuencia (un accidente). Comparamos el uso del teléfono móvil por los conductores en el momento estimado de un accidente con los mismos conductores durante otro momento conveniente. Como los conductores son sus propios controles, este diseño tiene en cuenta algunas características del conductor que podrían afectar la probabilidad de chocar, pero que no cambian en un periodo corto. Como es importante que los riesgos sean similares durante los periodos de control y los trayectos en que ocurre el accidente, comparamos el uso del

teléfono durante el intervalo de riesgo (tiempo inmediatamente anterior al accidente) con su uso durante los intervalos de control (periodos equivalentes durante los cuales los participantes estaban conduciendo pero no chocaron) en la semana previa²⁸.

Explicación

Aconsejamos presentar al comienzo los elementos clave del diseño del estudio, en la sección de métodos (o al final de la introducción), de manera que los lectores comprendan los fundamentos del estudio. Por ejemplo, se debe indicar que el estudio fue de cohortes, que siguió a los participantes durante un periodo de tiempo concreto, y describir al grupo de personas que integran la cohorte y su estatus de exposición. Igualmente, si la investigación usó un diseño de casos y controles se deben describir los casos y los controles, así como la fuente de la población. Si se trata de una encuesta hay que mencionar la población y el momento en que se tomó la muestra. Cuando se trate de una variante de los tres principales tipos de estudio, se requieren aclaraciones adicionales. Por ejemplo, respecto a un estudio de casos alternantes, que es una variante del diseño de casos y controles, en el ejemplo previo se ofreció una breve descripción de sus principios²⁸.

Recomendamos que los autores se abstengan de llamar simplemente a un estudio “prospectivo” o “retrospectivo”, porque estos términos están mal definidos²⁹. Un enfoque es entender “cohorte” y “prospectivo” como sinónimos, y reservar la palabra “retrospectivo” para estudios de casos y controles³⁰. Un segundo enfoque distingue los estudios de cohortes en prospectivos y retrospectivos, basándose en el momento en que surgió la idea del estudio y la temporalidad para recoger los datos³¹. Un tercer enfoque distingue estudios de casos y controles prospectivos y retrospectivos, dependiendo de si los datos acerca de la exposición de interés existían cuando se seleccionaron los casos³². Algunos recomiendan no utilizar estos términos³³ o adoptar las opciones “concurrente” e “histórica” para describir estudios de cohortes³⁴. En STROBE no usamos las palabras “prospectivo” y “retrospectivo”, ni opciones como “concurrente” e “histórica”. Recomendamos que, siempre que los autores utilicen estos términos, definan su significado. Más importante aún, recomendamos que los autores describan exactamente cómo y cuándo tuvo lugar la recogida de los datos.

La primera parte de la sección de métodos puede ser también el lugar apropiado para mencionar si el reporte es uno de varios otros derivados de un mismo estudio. Si un artículo está en línea con los propósitos originales del estudio, normalmente se indica refiriéndose a una publicación anterior y repitiendo de forma breve las características relevantes del estudio. Sin embargo, los propósitos del estudio también pueden evolucionar con el tiempo. Con frecuencia los investigadores usan datos con propósitos para los cuales no fueron planeados en principio, incluyendo, por ejemplo, estadísticas vitales que se elaboran principalmente con fines administrativos, preguntas de cuestionarios que originalmente sólo se incluyeron para hacer los datos más completos, o muestras de sangre tomadas con otros propósitos. Por ejemplo, el *Physicians' Health Study*, un ensayo controlado aleatorizado de ácido acetilsalicílico y carotenos, se utilizó posteriormente para demostrar que una mutación puntual en el gen del factor V se asociaba con un mayor riesgo de trombosis venosa, pero no con infarto de miocardio ni con enfermedad vascular cerebral³⁵. El uso secundario de datos existentes es una parte creativa de las investigaciones observacionales, y no necesariamente hace a los resultados menos creíbles o menos importantes. Sin embargo, repetir brevemente los intereses originales del estudio puede ayudar a los lectores a comprender el contexto de la investigación y las posibles limitaciones de los datos.

5. *Contexto: describa el marco, los lugares y las fechas relevantes, incluyendo los periodos de reclutamiento, exposición, seguimiento y recogida de datos*

Ejemplo

“El Estudio de Cohortes *Los Pasitos* reclutó mujeres embarazadas de las clínicas de mujeres, recién nacidos y niños en Socorro y San Elizario, en el condado de El Paso, Texas, y de las clínicas materno-infantiles del Instituto Mexicano de Seguridad Social en Ciudad Juárez, México, de abril de 1998 a octubre de 2000. Al inicio, previamente al nacimiento de los niños incluidos en la cohorte, se entrevistó a las madres respecto al ambiente familiar. Para el estudio nos propusimos realizar exámenes de seguimiento con intervalos de 6 meses, comenzando a los 6 meses de edad³⁶.”

Explicación

Los lectores necesitan información sobre el entorno y la ubicación para evaluar el contexto y la generalización de los resultados del estudio. Las exposiciones, los factores ambientales y los tratamientos pueden cambiar con el tiempo, así como también los métodos de estudio pueden evolucionar a lo largo del tiempo. Saber cuándo tuvo lugar un estudio y durante qué periodos fueron reclutados y seguidos los participantes, sitúa al estudio en un contexto histórico y es importante para la interpretación de los resultados.

La información acerca del marco incluye sitios o fuentes de reclutamiento (por ejemplo una lista electoral, una consulta externa, un registro de cáncer o centros de tercer nivel). La información sobre el lugar se puede referir a países, ciudades, hospitales o servicios clínicos donde tuvo lugar la investigación. Aconsejamos indicar las fechas en vez de sólo la extensión de los periodos de tiempo. Pueden existir diferentes fechas para determinar la exposición, la ocurrencia de la enfermedad, el reclutamiento, el comienzo y el fin del seguimiento, y la recogida de los datos. Cabe señalar que cerca del 80% de 132 artículos de revistas de oncología que usaron análisis de supervivencia incluían las fechas iniciales y finales para la acumulación de los pacientes, pero sólo el 24% indicaba la fecha en que terminó el seguimiento³⁷.

6. Participantes

6 (a). *Estudios de cohortes: proporcione los criterios de elegibilidad, así como las fuentes y el método de selección de los participantes. especifique los métodos de seguimiento*

Ejemplo

“Las participantes en el *Iowa Women's Health Study* fueron una muestra aleatoria de todas las mujeres de 55 a 69 años de edad obtenida de la lista de licencias de conducir del Estado de Iowa en 1985, que representaban aproximadamente el 94% de las mujeres de ese Estado en ese grupo de edad. (...) Los cuestionarios de seguimiento se enviaron por correo en octubre de 1987 y en agosto de 1989, para evaluar el estado vital y cambios de domicilio. (...) Se identificaron los cánceres incidentes, excepto el cáncer de piel tipo no melanoma, mediante el Registro Estatal de Salud de Iowa. (...) La cohorte del estudio fue emparejada con el Registro, con combinaciones de nombre, apellido y nombre de soltera, código postal, fecha de nacimiento y número de la seguridad social³⁸.”

6 (a). *Estudios de casos y controles: proporcione los criterios de elegibilidad, así como las fuentes y el proceso diagnóstico de los casos y el de selección de los controles. indique las razones para la elección de casos y controles*

Ejemplo

“Se identificaron los casos de melanoma cutáneo diagnosticados en 1999 y 2000 mediante del Registro de Cáncer de Iowa. (...)”

Los controles, también identificados por el mismo registro, fueron pacientes con cáncer colorrectal diagnosticados en el mismo periodo. Se seleccionaron controles con cáncer colorrectal debido a que son frecuentes y tienen una supervivencia relativamente larga, y porque la exposición a arsénico no se ha vinculado concluyentemente con la incidencia de cáncer colorrectal³⁹.

6 (a). *Estudios transversales: proporcione los criterios de elegibilidad, y las fuentes y los métodos de selección de los participantes*

Ejemplo

“Identificamos retrospectivamente pacientes con un diagnóstico principal de infarto de miocardio (código 410) según la Modificación Clínica de la Novena Revisión de la Clasificación Internacional de Enfermedades. De los códigos designados se descartaron diagnósticos excluyendo aquellos códigos cuyo quinto dígito fuera 2, que indica un episodio de atención subsecuente. (...) Se seleccionó una muestra aleatoria de la cohorte completa de Medicare con infarto de miocardio de febrero de 1994 a julio de 1995. (...) Para ser elegibles, los pacientes tenían que presentarse en el hospital transcurridos al menos 30 minutos con dolor de pecho, pero menos de 12 horas, y haber tenido una elevación del segmento ST de al menos 1 mm en dos derivaciones contiguas en el electrocardiograma inicial⁴⁰.”

Explicación

La descripción detallada de los participantes del estudio ayuda a los lectores a entender la aplicabilidad de los resultados. Los investigadores normalmente restringen la población de estudio mediante la definición de características clínicas, demográficas y otras, entre los participantes elegibles. Los criterios de elegibilidad típicos se relacionan con la edad, el sexo, el diagnóstico y las condiciones de comorbilidad. A pesar de su importancia, a menudo los criterios de elegibilidad no se describen adecuadamente. En un estudio sobre investigaciones observacionales sobre accidentes vasculares y cerebrales, 17 de 49 reportes (35%) no especificaron los criterios de elegibilidad⁹.

Los criterios de elegibilidad se pueden presentar como criterios de inclusión y exclusión, aunque esta distinción no siempre es necesaria o útil. No obstante, aconsejamos a los autores indicar todos los criterios de elegibilidad y describir el grupo del cual se seleccionó la población de estudio (por ejemplo la población general de una región o país), así como el método de reclutamiento (por ejemplo referencia o autoselección mediante anuncios).

Conocer los detalles sobre los procedimientos de seguimiento, incluyendo si éstos minimizaron la no respuesta y las pérdidas, así como si los procedimientos fueron similares para todos los participantes, permite juzgar la validez de los resultados. Por ejemplo, en un estudio que usó anticuerpos IgM para detectar infecciones agudas, los lectores necesitaban saber el intervalo entre los análisis de sangre para estos anticuerpos, de modo que pudieran estimar si se omitieron algunas infecciones debido a intervalos demasiado prolongados entre las pruebas⁴¹. En otros estudios en que los procedimientos de seguimiento difieren entre los grupos expuestos y no expuestos, los lectores podrían reconocer un sesgo sustancial debido a una evaluación desigual de los eventos a observar, a diferencias en la no respuesta o a pérdidas en el seguimiento⁴². Por lo tanto, aconsejamos que los investigadores describan los métodos usados para el seguimiento, si fueron los mismos para todos los participantes y cuán completa fue la medición de las variables (ver también punto 14).

En los estudios de casos y controles, la elección de los casos y de los controles es crucial para interpretar los resultados, y el método de selección tiene implicaciones importantes para la

validez del estudio. En general, los controles deben reflejar la población de donde surgen los casos. Se emplean varios métodos para muestrear a los controles, todos con ventajas y desventajas: para casos que surgen de una población general se emplean controles de muestras de listas de población, marcación de números telefónicos aleatorios, amigos o vecinos. Los controles de amigos o vecinos pueden presentar pareamiento intrínseco en la exposición¹⁷. Los controles con otras enfermedades pueden tener ventajas sobre los controles de base poblacional, en particular para casos hospitalarios, porque reflejan mejor la captación poblacional de un hospital, tienen una mayor comparabilidad recordatoria y son fáciles de reclutar. Sin embargo, pueden presentar problemas si la exposición de interés afecta al riesgo de desarrollar o ser hospitalizado por la condición de control^{43,44}. Para corregir este problema, con frecuencia se emplea una combinación de controles con las enfermedades probablemente menos asociadas a la exposición⁴⁵.

6 (b). *Estudios de cohortes: en los estudios pareados, proporcione los criterios para la formación de parejas y el número de participantes con y sin exposición*

Ejemplo

“Para cada paciente que inicialmente recibió estatinas usamos un emparejamiento basado en la propensión para identificar un control que no recibiera estatinas, según el siguiente protocolo. Primero se calculó la puntuación de propensión para cada paciente en toda la cohorte basándose en una larga lista de factores potencialmente relacionados con el uso de estatinas o con el riesgo de sepsis. Segundo, cada usuario de estatinas fue emparejado con un grupo más pequeño de no usuarios de estatinas por sexo, edad (± 1 año) y la fecha índice (± 3 meses). Tercero, seleccionamos el control con la puntuación de propensión más cercana (dentro del 0,2 de desviación estándar) para cada usuario de estatinas con una razón 1:1, y descartamos los controles restantes⁴⁶.”

6 (b). *Estudios de casos y controles: en los estudios emparejados, proporcione los criterios para la formación de las parejas y el número de controles por cada caso*

Ejemplo

“Nos propusimos seleccionar cinco controles por cada caso entre individuos de la población de estudio que no tuvieran diagnóstico de autismo ni otro trastorno generalizado del desarrollo (TGD) registrado en su expediente de atención primaria, y que estuvieran vivos y registrados como pacientes activos en la fecha del diagnóstico del TGD del caso. Los controles fueron pareados individualmente con los casos por año de nacimiento (hasta un año mayor o menor), sexo y centro de atención primaria. Para cada uno de los 300 casos se pudo identificar cinco controles que cumplieran todos los criterios de emparejamiento. Para los 994 restantes se excluyeron uno o más controles⁴⁷.”

Explicación

El pareamiento es mucho más común en estudios de casos y controles, pero en ocasiones los investigadores lo usan en los estudios de cohortes para hacer grupos comparables en el inicio del seguimiento. El pareamiento en los estudios de cohortes proporciona grupos directamente comparables por posibles confusores, y presenta menos complejidad que en los estudios de casos y controles. Por ejemplo, no es necesario tener en cuenta el pareamiento para la estimación del riesgo relativo. Ya que en los estudios de cohortes el pareamiento puede aumentar la precisión

Recuadro 2–Pareamiento en estudios de casos y controles.

En cualquier estudio de casos y controles es necesario tomar decisiones razonables respecto a si usar pareamiento, y de ser así qué variables parear, el método preciso para hacerlo y el apropiado análisis estadístico. El hecho de no parear por ninguna variable puede significar que la distribución de algunos posibles confusores (p.ej., edad, sexo) sea radicalmente diferente entre casos y controles. Si bien en el análisis puede ajustarse por estas variables, es probable que haya una pérdida importante de la eficiencia estadística.

El uso del pareamiento en los estudios de casos y controles, así como su interpretación, están llenos de dificultades, especialmente si se para por varios factores de riesgo, algunos de los cuales pueden estar asociados a la exposición de interés^{50,51}. Por ejemplo, en un estudio de casos y controles sobre infarto de miocardio y anticonceptivos orales, anidado en una gran base de datos farmacoepidemiológica, con información sobre miles de mujeres disponibles como posibles controles, los investigadores pueden estar tentados a elegir controles pareados con similares factores de riesgo por cada caso de infarto de miocardio. Un objetivo es ajustar por factores que pueden influir en la prescripción de anticonceptivos orales, y de esta manera controlar la confusión por indicación. Sin embargo, el resultado será un grupo control que deja de ser representativo del uso de anticonceptivos orales en la población base: los controles serán más viejos que la población base debido a que los pacientes con infarto de miocardio tienden a ser más viejos. Esto tiene varias implicaciones. Un análisis crudo de los datos puede producir *odds ratios* normalmente sesgadas hacia la unidad si el factor de pareamiento está asociado con la exposición. La solución es realizar un análisis pareado o estratificado (véase el punto 12b). Asimismo, debido a que el grupo de controles pareados dejó de ser representativo de la población, la distribución de la exposición entre los controles no puede usarse para estimar la fracción atribuible poblacional (véase el **Recuadro 7**)⁵². Además, ya no se puede estudiar el efecto del factor de pareamiento, y la búsqueda de controles bien pareados puede ser engorrosa, lo que hace preferible un diseño no pareado debido a que los controles no pareados serían más sencillos de obtener y el grupo control puede ser más grande. El sobrepareamiento es otro problema que puede reducir la eficiencia de los estudios de casos y controles pareados, y en algunas situaciones puede introducir sesgos. La información se pierde y el poder del estudio se reduce si la variable de pareamiento está estrechamente asociada con la exposición. Por lo tanto, muchos individuos dentro de los mismos grupos pareados tenderán a tener idénticas o similares exposiciones, y por lo tanto no contribuyen con información relevante. El pareamiento irremediablemente introducirá sesgos si la variable de pareamiento no es un confusor sino que se encuentra en la ruta causal entre la exposición y la enfermedad. Por ejemplo, la fertilización *in vitro* está asociada con un riesgo elevado de muerte perinatal, debido a un aumento de los nacimientos múltiples y del bajo peso al nacer⁵³. Parear por pluralidad o peso al nacer sesgará los resultados hacia la nulidad, y esto no se puede remediar en el análisis.

El pareamiento es intuitivamente atractivo, pero la complejidad que implica ha llevado a los metodólogos a advertir contra la sistematización del pareamiento en los estudios de casos y controles. Recomiendan hacer una consideración cuidadosa y juiciosa de cada posible factor de pareamiento, reconociendo que pueden ser determinados y usados como una variable de ajuste en lugar de parear por ellos. En consecuencia, ha habido una reducción en el número de factores de pareamiento utilizados y un uso creciente del pareamiento por frecuencia, lo que evita algunos de los problemas comentados anteriormente, así como más estudios de casos y controles no pareados en absoluto⁵⁴. El pareamiento continúa siendo lo más deseable, e incluso

necesario, cuando la distribución de los confusores (p.ej., la edad) puede diferir radicalmente entre los grupos de comparación no pareados^{48,49}.

estadística, los investigadores pueden permitir el pareamiento en sus análisis y además obtener intervalos de confianza más estrechos.

Los estudios de casos y controles pareados se realizan para incrementar la eficiencia del estudio y asegurar la similitud en la distribución de las variables entre casos y controles, en particular la distribución de las variables potencialmente confusoras^{48,49}. Puesto que el pareamiento se puede hacer de varias formas, con uno o más controles por caso, se deben describir la justificación para la selección de variables de emparejamiento y los detalles de los métodos utilizados. Las formas de emparejamiento comúnmente usadas son por frecuencia (también llamado “pareamiento o emparejamiento de grupo”) e individual. En el pareamiento por frecuencia, los investigadores eligen controles de manera que la distribución de las variables de pareamiento sea idéntica o similar que la de los casos. El pareamiento individual implica parear uno o varios controles a cada caso. Aunque intuitivamente es atractivo y algunas veces útil, el pareamiento en los estudios de casos y controles tiene ciertas desventajas, no siempre es apropiado y debe tenerse en cuenta en el análisis (véase el **Recuadro 2**).

Incluso procedimientos de pareamiento simples en apariencia pueden ser mal comunicados. Por ejemplo, los autores pueden establecer que los controles se parearon a los casos “dentro de un periodo de 5 años” o “usando intervalos de edad de 5 años”. ¿Esto significa que, si un caso tenía 54 años de edad, el respectivo control necesitaba estar en el intervalo de los 5 años, de 50 a 54, o con una edad de 49 a 59, que está dentro de los 5 años a partir de la edad de 54? Si se elige un intervalo de edad amplio (por ejemplo 10 años) hay riesgo de una confusión residual por edad (véase también el **Recuadro 4**), ya que, por ejemplo, los controles pueden ser en promedio más jóvenes que los casos.

7. Variables: defina claramente todas las variables, de respuesta, exposiciones, predictoras, confusoras y modificadoras del efecto. si procede, proporcione los criterios diagnósticos

Ejemplo

“Sólo se incluyeron en el análisis malformaciones congénitas importantes. Se excluyeron las anomalías secundarias según la lista de exclusión de las Anomalías Congénitas del Registro Europeo (EUROCAT). Si un niño tenía más de una malformación congénita importante en un órgano, estas malformaciones fueron tratadas como un evento resultado en el análisis por sistema de órganos. (...) En el análisis estadístico, los factores considerados como posibles confusores fueron la edad materna en el momento del parto y el número de partos previos. Los factores considerados como posibles modificadores del efecto fueron la edad materna en el momento del reembolso del medicamento antiepiléptico y la edad materna en el parto”⁵⁵.

Explicación

Los autores deben definir todas las variables consideradas e incluidas en el análisis, de respuesta, exposiciones, predictoras y posibles confusoras y modificadoras del efecto. Las enfermedades resultado requieren una descripción detallada y adecuada de los criterios diagnósticos. Esto es de aplicación a los criterios para casos en un estudio de casos y controles, a los eventos de enfermedad durante el seguimiento en un estudio de cohortes, y a

la enfermedad prevalente en un estudio transversal. Son particularmente importantes las definiciones claras y los pasos tomados para adherirse a ellas para cualquier enfermedad de interés principal en el estudio.

Para algunos estudios, “determinante” y “predictor” pueden ser términos apropiados para variables de exposición, y las de resultado pueden llamarse desenlaces. En modelos multivariados, algunas veces los autores usan “variable dependiente” para un resultado y “variable independiente” o “variable explicatoria” para la exposición y las variables confusoras. Esto último no es necesario, ya que no distingue la exposición de los confusores.

Si se han medido e incluido muchas variables en el análisis exploratorio en una fase temprana de la investigación, considere proporcionar una lista con detalles sobre cada variable en un apéndice, tabla adicional o publicación separada. El *International Journal of Epidemiology* ha creado recientemente una nueva sección con “perfiles de cohortes”, que incluye información detallada sobre lo que se midió en diferentes momentos en el tiempo en estudios particulares^{56,57}. Finalmente, recomendamos a los autores que declaren todas las “variables candidatas” consideradas para el análisis estadístico, en lugar de indicar selectivamente sólo las incluidas en los modelos finales (ver también el punto 16a)^{58,59}.

8. Fuentes de datos/medidas: para cada variable de interés, indique las fuentes de datos y los detalles de los métodos de valoración (medida), si hubiera más de un grupo, especifique la comparabilidad de los procesos de medida

Ejemplo 1

“Se calculó principalmente el consumo de cafeína total usando las fuentes de composición de alimentos del Departamento de Agricultura de Estados Unidos. En estos cálculos se asumió que el contenido de cafeína era de 137 mg por taza de café, 47 mg por taza de té, 46 mg por lata o botella de bebida de cola, y 7 mg por ración de chocolate. Este método de medida del consumo (cafeína) mostró ser válido tanto en el estudio de la cohorte NHS I como en un estudio de una cohorte similar de profesionales de la salud varones. (...) Se halló que el diagnóstico auto-reportado de hipertensión era fiable en la cohorte NHS I”⁶⁰.

Ejemplo 2

“Las muestras pertenecientes a los casos y los controles pareados se analizaron siempre juntas en la misma serie, y el personal del laboratorio no podía distinguir entre casos y controles”⁶¹.

Explicación

La forma en que se miden la exposición, los confusores y los resultados afecta la fiabilidad y la validez de un estudio. El error de medida y la mala clasificación de la exposición o el resultado pueden hacer más difícil detectar una relación causa-efecto, o pueden producir relaciones falsas. El error en la medida de posibles confusores puede incrementar el riesgo de confusión residual^{62,63}. Por lo tanto, es útil que los autores comuniquen los hallazgos de cualquier estudio de validez o fiabilidad de las evaluaciones o medidas, incluyendo detalles del estándar de referencia que se utilizó. En lugar de simplemente citar los estudios de validación (como en el primer ejemplo), aconsejamos que los autores proporcionen la validez y la fiabilidad estimadas que puedan utilizarse para un ajuste del error de medida o un análisis de sensibilidad (ver puntos 12e y 17).

Además, es importante saber si los grupos que se están comparando son diferentes respecto a la forma en que se recogieron los datos. Esto puede ser importante para análisis de

laboratorio (como en el segundo ejemplo) y otras situaciones. Por ejemplo, si un entrevistador primero entrevista a todos los casos y después a los controles, o viceversa, es posible un sesgo debido a una curva de aprendizaje; soluciones como aleatorizar el orden de las entrevistas pueden evitar estos problemas. También pueden aparecer sesgos de información si a los grupos de comparación no se les aplican las mismas pruebas diagnósticas, o si un grupo recibe más pruebas del mismo tipo que el otro (ver también punto 9).

9. Sesgos: especifique todas las medidas adoptadas para afrontar posibles fuentes de sesgo

Ejemplo 1

“En la mayoría de los estudios de casos y controles de suicidio el grupo control comprende individuos vivos, pero decidimos tener un grupo control de personas que murieron por otras causas. (...) Con un grupo control de individuos muertos, las fuentes de información utilizadas para evaluar los factores de riesgo son informantes que han experimentado recientemente la muerte de un familiar o de un conocido cercano, y por consiguiente son más comparables con las fuentes de información del grupo de suicidas que si se usaran controles vivos”⁶⁴.

Ejemplo 2

“El sesgo de descubrimiento podría influir en la asociación entre diabetes mellitus de tipo 2 (DMT2) y glaucoma primario de ángulo abierto (GPAA) si las mujeres con DMT2 estuvieran bajo una vigilancia oftálmica más estricta que las mujeres sin esta condición. Comparamos el número promedio de exploraciones oculares referido por las mujeres con y sin diabetes. También recalculamos el riesgo relativo de GPAA con un control adicional para covariables asociadas a una vigilancia ocular más cuidadosa (autorreporte de cataratas, degeneración macular, número de exploraciones oculares y número de exploraciones físicas)”⁶⁵.

Explicación

Los estudios con sesgos producen resultados que difieren sistemáticamente de la verdad (véase también el **Recuadro 3**). Es importante para el lector saber qué medidas se tomaron durante el desarrollo del estudio para reducir los posibles sesgos. Idealmente, cuando planean el estudio los investigadores consideran con sumo cuidado las posibles fuentes de sesgo. En la fase del informe, recomendamos que evalúen siempre la probabilidad de sesgos relevantes. Específicamente debe considerarse la dirección y la magnitud del sesgo, y si es posible estimarlo. Por ejemplo, en los estudios de casos y controles la información puede sesgarse, pero puede reducirse seleccionando un grupo control apropiado, como en el primer ejemplo⁶⁴. En el segundo ejemplo⁶⁵, las diferencias en la vigilancia médica de las participantes era un problema. Por consiguiente, los autores proporcionan más detalles sobre los datos adicionales que recogieron para solucionar este problema. Se debe describir si los investigadores han establecido programas de control de calidad para la recogida de los datos, para contrarrestar un posible “cambio” en las medidas de las variables en estudios longitudinales o para mantener una mínima variabilidad cuando se utilizan observadores múltiples.

Desafortunadamente, a menudo los autores no refieren los sesgos importantes al comunicar sus resultados. De 43 estudios de casos y controles y de cohortes publicados entre 1990 y 1994 que investigaban el riesgo de segundos cánceres en pacientes con historia de cáncer, el sesgo de vigilancia médica se mencionaba sólo en cinco artículos⁶⁶. Una encuesta de artículos sobre investigación en salud mental publicados en 1998 en tres revistas psiquiátricas encontró que sólo el 13% de 392 artículos

Recuadro 3–Sesgos.

Un sesgo es una desviación sistemática del valor verdadero del resultado de un estudio. Típicamente se introduce durante el diseño o el desarrollo de un estudio, y después ya no puede ser solucionado. Sesgo y confusor no son sinónimos. Un sesgo es el resultado de un defecto en la obtención de la información o la selección de los sujetos, de manera que se encuentra una asociación errónea. La confusión produce relaciones objetivamente ciertas, pero que no pueden ser interpretadas como causales debido a que un factor subyacente, no considerado, está asociado tanto con la exposición como con el resultado (véase el **Recuadro 5**). Asimismo, es necesario distinguir sesgo de error aleatorio, que es una desviación de un valor verdadero ocasionada por fluctuaciones estadísticas (en cualquier dirección) en la medida de los datos. Se han descrito muchas posibles fuentes de sesgo usando una variedad de términos^{68,69}. Nosotros encontramos útiles dos categorías simples: sesgo de información y sesgo de selección.

El sesgo de información ocurre cuando diferencias sistemáticas en la completitud o la exactitud de los datos conducen a una mala clasificación diferencial de los individuos con respecto a las exposiciones o a los eventos resultado. Por ejemplo, si las mujeres diabéticas son sometidas a exámenes oculares más regulares y minuciosos, el hallazgo de glaucoma será más frecuente que en las mujeres sin diabetes (véase el punto 9)⁶⁵. Los pacientes que toman un medicamento que causa molestias estomacales inespecíficas pueden someterse a una gastroscopia con más frecuencia y tener más úlceras detectadas que aquellos que no reciben el medicamento, aun cuando éste no ocasione más úlceras. Este tipo de sesgo de información también se conoce como “sesgo de detección” o “sesgo de vigilancia médica”. Una forma de evaluar su influencia es medir la intensidad de la vigilancia médica en los diferentes grupos de estudio, y ajustar por ésta en el análisis estadístico. En los estudios de casos y controles, el sesgo de información ocurre si los casos recuerdan las exposiciones pasadas con más precisión que los controles, o si tienen mejor disposición para relatarlas (también llamado “sesgo de recordatorio”). El “sesgo del entrevistador” puede ocurrir si los entrevistadores conocen la hipótesis de estudio y, de manera consciente o inconsciente, recogen información selectiva⁷⁰. Por tanto, resulta valioso “cegar” a los participantes del estudio y a los investigadores.

El sesgo de selección puede introducirse en los estudios de casos y controles si la probabilidad de incluir casos o controles está asociada con la exposición. Por ejemplo, un médico que recluta pacientes para un estudio sobre trombosis venosa profunda es posible que diagnostique esta enfermedad en una mujer que tiene molestias en las piernas y toma anticonceptivos orales. Sin embargo, probablemente no la diagnostique en una mujer con molestias similares pero que no toma tales medicamentos. Este sesgo puede contrarrestarse usando casos y controles que sean remitidos de la misma forma al servicio donde se realice el diagnóstico⁷¹. Del mismo modo, el uso de registros de enfermedades puede introducir sesgo de selección: si se conoce una posible relación entre una exposición y una enfermedad, los casos pueden resultar más propensos a ser registrados si han estado expuestos al agente causal en estudio⁷². El “sesgo de respuesta” es otro tipo de sesgo de selección que ocurre si las diferencias en las características entre los que responden y los que no aceptan participar en un estudio afectan la estimación de las prevalencias, las incidencias y, en algunas circunstancias, las asociaciones. En general, el sesgo de selección afecta la validez interna de un estudio. Esto es diferente de los problemas que pueden presentarse con la selección de los participantes para un estudio en general, lo cual afecta la validez externa más que la interna (véase también el punto 21).

mencionaban el sesgo de respuesta⁶⁷. Un trabajo sobre estudios de cohortes sobre accidentes vasculares cerebrales encontró que 14 de 49 (28%) artículos publicados entre 1999 y 2003 indicaron un posible sesgo de selección en el reclutamiento de los participantes del estudio, y 35 (71%) mencionaban la posibilidad de que algún tipo de sesgo pudiera afectar sus resultados⁵.

10. Tamaño muestral: explique cómo se determinó el tamaño muestral**Ejemplo 1**

“El número de casos en el área durante el periodo de estudio determinó el tamaño de la muestra”⁷³.

Ejemplo 2

“Una encuesta sobre depresión posparto en la región había identificado una prevalencia del 19,8%. Asumiendo que la depresión en madres de niños con peso normal es del 20% y una *odds ratio* de 3 para depresión en madres de niños con desnutrición, necesitábamos 72 conjuntos de caso y control (un caso por un control) con un poder del 80% y una significación del 5%”⁷⁴.

Explicación

Un estudio debe ser lo bastante grande como para obtener una estimación concreta con un intervalo de confianza suficientemente estrecho que conteste de manera significativa la pregunta de investigación. Se necesitan muestras grandes para distinguir una asociación pequeña. Los estudios pequeños a menudo proporcionan valiosa información; sin embargo, los intervalos de confianza amplios contribuyen poco al conocimiento actual, comparados con los estudios que proporcionan estimadores con intervalos de confianza más pequeños. También se publican con más frecuencia estudios pequeños que muestran asociaciones “estadísticamente significativas” o “interesantes” que estudios pequeños que no tienen resultados “significativos”. Aunque tales estudios pueden proporcionar una señal temprana acerca de una posible asociación, debe informarse a los lectores de sus posibles debilidades.

La importancia de determinar el tamaño de la muestra en los estudios observacionales depende del contexto. Si el análisis se realiza con datos que ya estaban disponibles para otros propósitos, la pregunta principal es si el análisis de los datos producirá resultados con la suficiente precisión estadística para contribuir sustancialmente a la literatura, y las consideraciones del tamaño de la muestra serán relativamente secundarias. El cálculo formal, a priori, del tamaño de la muestra puede ser útil cuando se planifica un nuevo estudio^{75,76}. Tales cálculos se asocian con una mayor incertidumbre que la que implica la sola cifra que generalmente se produce. Por ejemplo, las estimaciones de la tasa del evento de interés u otras suposiciones centrales para los cálculos normalmente son imprecisas, si no meras conjeturas⁷⁷. La precisión obtenida en el análisis final no puede determinarse de antemano, ya que se reducirá por la inclusión de las variables de confusión en los análisis multivariados⁷⁸, por el grado de precisión con que se puedan medir las variables importantes⁷⁹ o por la exclusión de algunos individuos.

Pocos estudios epidemiológicos explican las decisiones sobre el tamaño de la muestra^{4,5}. Instamos a los investigadores para que describan el cálculo formal pertinente del tamaño de la muestra. En otras situaciones se deben indicar las consideraciones que determinaron el tamaño del estudio (por ejemplo, una muestra disponible fija, como en el primer ejemplo de arriba). Se debe mencionar si el estudio se concluyó antes, cuando se alcanzó la significación estadística. No importune a los lectores con justificaciones *post hoc* del tamaño del estudio o con cálculos de poder retrospectivos⁷⁷. Desde el punto de vista del lector, los intervalos

de confianza indican la precisión estadística que finalmente se obtuvo. Hay que tener en cuenta que los intervalos de confianza sólo reflejan la incertidumbre estadística, no toda la incertidumbre que puede estar presente en un estudio (véase el punto 20).

11. Variables cuantitativas: explique cómo se trataron las variables cuantitativas en el análisis. si procede, explique qué grupos se definieron y por qué

Ejemplo

“Se considera que los pacientes con una puntuación menor de 8 en la Escala de Coma de Glasgow (GSC) presentan un daño grave. Una puntuación GCS de 9 o más indica una lesión cerebral menos grave. Investigamos la asociación de estos dos valores GCS con la ocurrencia de muerte dentro de los 12 meses siguientes a la lesión”⁸⁰.

Explicación

Los investigadores toman decisiones respecto a cómo recoger y analizar los datos cuantitativos sobre las exposiciones, los modificadores del efecto y los confusores. Por ejemplo, pueden agrupar una exposición continua para crear una nueva variable categórica (véase el Recuadro 4). Las decisiones respecto al agrupamiento pueden tener consecuencias importantes para los análisis posteriores^{81,82}. Recomendamos que los autores expliquen por qué y cómo se agruparon los datos cuantitativos, incluyendo el número de categorías, los puntos de corte y la media o la mediana de la categoría. Siempre que los datos se presenten en forma de tabla debe proporcionarse el número de casos, de controles, de personas en riesgo, de personas-tiempo, el riesgo, etc., para cada categoría. Las tablas no deben consistir solamente en estimaciones de la medida del efecto o resultados del ajuste del modelo.

Los investigadores podrían modelar una exposición como continua con el fin de conservar toda la información. Al tomar esta decisión, se necesita considerar la naturaleza de la relación de la exposición con el resultado. Así como puede ser incorrecto asumir automáticamente una relación lineal, también deben investigarse posibles desviaciones de la linealidad. Los autores podrían mencionar los modelos alternativos que exploraron durante el análisis (por ejemplo, usando una transformación logarítmica, términos cuadráticos o funciones articuladoras). Existen varios métodos para ajustar una relación no lineal entre la exposición y el resultado^{82–84}. También puede ser informativo presentar ambos análisis, el continuo y el agrupado, para una exposición cuantitativa de interés primario.

En un estudio reciente, dos tercios de las publicaciones epidemiológicas incluyeron variables de exposición cuantitativas⁴. En 42 de 50 artículos (84%), la exposición se agrupó en categorías ordenadas, pero sin justificar esta decisión. Quince artículos usaron asociaciones lineales para modelar la exposición continua, pero sólo dos verificaron la linealidad. En otro estudio sobre literatura de psicología⁸⁵, la dicotomización sólo se justificaba en 22 de 110 artículos (20%).

12. Métodos estadísticos

12 (a). especifique todos los métodos estadísticos, incluidos los empleados para controlar los factores de confusión

Ejemplo

“Se calculó el riesgo relativo ajustado usando la técnica de Mantel-Haenzel para evaluar la presencia de confusión por edad o sexo en los grupos de comparación. Se estimó el intervalo de confianza (IC) del 95% del riesgo relativo ajustado, usando la varianza de acuerdo con Greenland y Robins, y Robins et al”⁹³.

Recuadro 4–Agrupación.

Los datos continuos pueden agruparse por diferentes razones⁸⁶. Cuando se recogen datos es mejor buscar una variable ordinal más que intentar obtener una medida continua artificialmente precisa para una exposición que se basa en el recuerdo de varios años atrás. Las categorías también pueden ser útiles con propósitos de presentación, por ejemplo para presentar todas las variables en un estilo similar o para mostrar una relación dosis-respuesta.

También puede ser útil agrupar para simplificar el análisis, por ejemplo para evitar una suposición de linealidad. Sin embargo, al agrupar se pierde información y se puede reducir el poder estadístico⁸⁷, en especial cuando se usa la dicotomización^{82,85,88}. Si se agrupa un confusor continuo puede aparecer confusión residual, mediante la cual parte del efecto confusor de la variable permanece sin ajustar (véase el Recuadro 5)^{62,89}. Aumentar el número de categorías puede disminuir la pérdida de poder y la confusión residual, y es especialmente apropiado en estudios grandes. Los estudios pequeños pueden usar pocos grupos debido a la limitación en los números.

Para fines prácticos, los investigadores pueden elegir puntos de corte como criterios de agrupación basándose en valores de uso común que son relevantes para el diagnóstico o el pronóstico. También pueden hacerlo por medio de criterios estadísticos. Pueden elegir igual número de individuos en cada grupo usando percentiles⁹⁰. Por otro lado, se puede ganar más claridad en la asociación con el evento resultante eligiendo grupos extremos y manteniendo el grupo o grupos intermedios más grandes que los grupos de los extremos⁹¹. En los estudios de casos y controles es preferible derivar una distribución del grupo control debido a su intento por representar a la población base. Debe informarse a los lectores si los puntos de corte fueron seleccionados *post hoc* tras varias alternativas. En particular, si los puntos de corte fueron elegidos para minimizar un valor de *p*, la verdadera fuerza de una asociación puede estar exagerada⁸¹.

Cuando se analizan las variables agrupadas es importante reconocer su naturaleza continua subyacente. Por ejemplo, puede investigarse una posible tendencia en el riesgo a través de grupos ordenados. Un abordaje común es modelar el rango de los grupos como una variable continua. Tal linealidad en las puntuaciones de los grupos aproximará una asociación lineal real si los grupos están igualmente espaciados (por ejemplo, grupos con amplitud de 10 años de edad), pero no de otra forma. Il'yasova et al⁹² recomiendan publicar las estimaciones tanto categórica como continua del efecto, con sus errores estándar, a fin de facilitar el metaanálisis, así como proporcionar información intrínsecamente valiosa sobre dosis-respuesta. Un análisis puede informar sobre el otro y ninguno está libre de asunciones. Con frecuencia los autores ignoran el orden y consideran las estimaciones (y los valores de *p*) por separado para cada categoría, comparándolas con la de referencia. Esto puede ser útil para fines descriptivos, pero puede no serlo para detectar una tendencia real en el riesgo a través de los grupos. Si se observa una tendencia, un intervalo de confianza para una pendiente podría indicar la fuerza de la observación.

Explicación

En general no hay un único análisis estadístico correcto, sino que más bien existen varias posibilidades que pueden atender a la misma cuestión haciendo diferentes suposiciones. Independientemente de esto, en el protocolo de investigación los investigadores deben determinar por anticipado al menos los análisis para los objetivos primarios del estudio. Con frecuencia se requieren

Recuadro 5–Confusión.

Confusión significa literalmente “confusión de efectos”. Un estudio puede parecer estar mostrando una asociación o una no asociación entre una exposición y el riesgo de una enfermedad, y en realidad la aparente asociación o su falta se debe a otro factor que determina la ocurrencia de la enfermedad, pero que también está asociado con la exposición. Ese otro factor se conoce como “factor de confusión” o “confusor”. Por tanto, la confusión da lugar a una evaluación errónea de la posible asociación “causal” de una exposición. Por ejemplo, si a las mujeres que se acercan a la edad madura y desarrollan hipertensión se les prescribe con menor frecuencia anticonceptivos orales, una comparación simple de la frecuencia de enfermedad cardiovascular entre aquellas que usan anticonceptivos y las que no, puede dar la impresión equivocada de que los anticonceptivos protegen contra la enfermedad cardíaca.

Los investigadores deben pensar de antemano en los posibles factores de confusión. Esto ayudará a elegir el diseño del estudio y permitirá una recogida de datos adecuada, ya que al identificar los confusores será más sencillo guiar la búsqueda de información detallada. Pueden usarse la restricción o el pareamiento. En el ejemplo anterior, el estudio podría restringirse a mujeres que no tengan el confusor, la hipertensión. También se podría parear por la presión arterial, aunque no necesariamente es deseable (véase el **Recuadro 2**). En la fase de análisis, los investigadores pueden usar un análisis estratificado o un análisis multivariado para reducir el efecto de los confusores. La estratificación consiste en dividir los datos en estratos según el confusor (por ejemplo, estratificar por presión arterial), evaluar la estimación de la asociación dentro de cada estrato y calcular la estimación combinada de la asociación así como un promedio ponderado de todos los estratos. El análisis multivariado logra el mismo resultado, pero permite considerar otras variables simultáneamente. Es más flexible, pero puede implicar suposiciones adicionales acerca de la forma matemática de la relación entre la exposición y la enfermedad.

En los estudios observacionales es de suma importancia tener en cuenta los confusores; sin embargo, los lectores no deben asumir que el análisis ajustado por confusores establece la “parte causal” de una asociación. Los resultados todavía pueden estar distorsionados por confusión residual (la confusión que permanece después de intentos no exitosos de controlar por confusores¹⁰²), por error de muestreo aleatorio, por sesgo de selección y por sesgo de información (véase el **Recuadro 3**).

proceso. Es útil informar a los lectores de si una o dos covariables son causa de gran parte de la confusión aparente en un análisis de datos. También se deben describir otros análisis estadísticos, como los procedimientos de imputación, transformación de datos y cálculos de riesgos atribuibles. Se deben indicar las referencias de procedimientos novedosos o poco comunes, así como el *software* estadístico utilizado. Recomendamos que los métodos estadísticos se describan “con suficiente detalle para permitir al lector con conocimientos y acceso a los datos originales verificar los resultados comunicados”¹⁰⁰.

En un estudio empírico, sólo 93 de 169 artículos (55%) que indicaban ajuste por confusión establecían claramente cómo se introdujeron las variables continuas y categóricas en el modelo estadístico¹⁰¹. Otro estudio encontró que en la mayoría de los 67 artículos en que los análisis estadísticos se ajustaban por confusores no estaba claro cómo se habían escogido éstos⁴.

12 (b). *Especifique todos los métodos utilizados para analizar subgrupos e interacciones*

Ejemplo

“Se exploraron las diferencias por sexo en la susceptibilidad a tres factores de riesgo relacionados con el estilo de vida, mediante la comprobación de la interacción biológica según Rothman: se redefinió una nueva variable compuesta por cuatro categorías para la variable sexo y una exposición de interés dicotómica (a^-b^- , a^-b^+ , a^+b^- y a^+b^+), donde a^- y b^- indican ausencia de la exposición. Se calculó el RR para cada categoría después de ajustar por la edad. Se definió efecto de interacción por la aditividad de los efectos absolutos, y se calculó el exceso de RR causado por la interacción (RERI):

$$RERI = RR(a^+b^+) - RR(a^-b^+) - RR(a^+b^-) - 1$$

donde $RR(a^+b^+)$ es el RR de aquellos expuestos a ambos factores y $RR(a^-b^-)$ es la categoría de referencia ($RR = 1,0$). Se calcularon los intervalos de confianza del 95% como proponen Hosmer y Lemeshow. Un RERI de 0 significa que no hay interacción”¹⁰³.

Explicación

Como se comenta con detalle en el punto 17, el uso y el valor del análisis restringido de subgrupos de la población de estudio es objeto de debate^{4,104}. No obstante, es frecuente que se hagan análisis de subgrupos⁴. Los lectores necesitan saber qué análisis de subgrupos se planificaron de antemano y cuáles surgieron mientras se analizaban los datos. También es importante explicar qué métodos se usaron para examinar si los efectos o asociaciones eran diferentes entre los grupos (véase el punto 17).

La interacción se refiere a la situación en la cual un factor modifica el efecto de otro (de ahí la expresión “modificación del efecto”). La acción conjunta de dos factores se puede caracterizar de dos maneras: en una escala aditiva en términos de diferencias de riesgo, o en una escala multiplicativa en términos de riesgo relativo (véase el **Recuadro 8**). Muchos autores y lectores pueden tener preferencias sobre la manera en que deben analizarse las interacciones, e incluso pueden estar interesados en saber hasta qué punto el efecto conjunto de las exposiciones difiere de los efectos separados. Hay acuerdo general en que la escala aditiva, que usa los riesgos absolutos, es más apropiada para la toma de decisiones en salud pública y en la práctica clínica¹⁰⁵. Cualquier punto de vista que se adopte debe presentarse claramente al lector, como se hace en el ejemplo anterior¹⁰³. Lo más informativo puede ser una presentación que muestre los efectos de ambas exposiciones por separado y su efecto conjunto, cada una respecto a la no exposición. En el punto 17 se presenta el ejemplo de la

análisis adicionales motivados por los datos, ya sea en lugar o además de los que originalmente se previeron. Al comunicar un estudio, los autores deben mencionar si los análisis particulares se debieron a la inspección de los datos. Aunque la distinción entre análisis preestablecido y exploratorio a veces puede ser ambigua, los autores deben aclarar las razones para los análisis particulares.

Si los grupos de comparación no son similares respecto a algunas características, el ajuste se debe hacer para las posibles variables confusoras mediante estratificación o regresión multivariada (véase el **Recuadro 5**)⁹⁴. Con frecuencia el diseño del estudio determina la selección del tipo de análisis de regresión. Por ejemplo, la regresión de riesgo proporcional de Cox normalmente se usa en estudios de cohortes⁹⁵, mientras que la regresión logística es a menudo el método de elección en los estudios de casos y controles^{96,97}. Los analistas deben describir totalmente los procedimientos específicos para la selección de las variables, y no sólo presentar los resultados del modelo final^{98,99}. Si se comparan modelos para determinar una lista de posibles confusores que se incluirán en un modelo final, también debe describirse este

interacción y en el **Recuadro 8** se explican los cálculos sobre las diferentes escalas.

12 (c). *Explique el tratamiento de los datos ausentes (missing data)*

Ejemplo

“Nuestros procedimientos de análisis de datos ausentes se basaron en suposiciones de datos ausentes al azar. Usamos el método de imputación múltiple multivariada por ecuaciones encadenadas en STATA. Analizamos independientemente 10 copias de los datos, cada una con valores ausentes adecuadamente imputados en los análisis de regresión logística. Promediamos las estimaciones de las variables para dar una sola media estimada y los errores estándar ajustados según las reglas de Rubin”¹⁰⁶.

Explicación

Los datos ausentes son comunes en la investigación observacional. Las encuestas enviadas por correo a los participantes no siempre están contestadas por completo, los participantes pueden no asistir a todas las visitas subsecuentes, y las fuentes habituales de datos, así como las bases de datos clínicos, a menudo están incompletas. A pesar de su ubicuidad e importancia, pocos artículos informan con detalle sobre el problema de los datos ausentes^{5,107}. Los investigadores pueden usar cualquiera de varios enfoques para manejar los datos ausentes. En el **Recuadro 6** describimos algunas ventajas y limitaciones de diversos planteamientos. Aconsejamos que los autores informen del número de valores ausentes para cada variable de interés (exposición, resultado, confusores) y para cada paso en el análisis. Si es posible, los autores deben dar las razones de los valores ausentes, e indicar cuántos individuos se excluyeron debido a datos ausentes al describir el flujo de participantes a lo largo del estudio (véase también el punto 13). Para los análisis que tengan en cuenta datos ausentes, los autores deben describir la naturaleza del análisis (por ejemplo, imputación múltiple) y las suposiciones que se hicieron (por ejemplo, pérdida al azar; véase el **Recuadro 6**).

12 (d). *Estudios de cohortes: si procede, explique cómo se afrontan las pérdidas en el seguimiento*

Ejemplo

“En programas de tratamiento con seguimiento activo, aquellos que se perdieron en el seguimiento presentaron en la determinación basal un recuento similar de células CD4 que aquellos con seguimiento a un año (mediana de 115 células por μl y 123 células por μl), mientras que los pacientes que se perdieron en el seguimiento en programas sin procedimientos de seguimiento activo tuvieron recuentos de CD4 considerablemente más bajos que los que se siguieron (mediana de 64 células por μl y 123 células por μl). (...) Los programas de tratamiento con seguimiento pasivo se excluyeron de los análisis subsecuentes”¹¹⁶.

Explicación

Los estudios de cohortes se analizan usando métodos de tablas de vida u otros enfoques basados en las personas-tiempo de seguimiento y en el tiempo para desarrollar la enfermedad de interés. Se asume que en los individuos que permanecen libres de la enfermedad al final del periodo de observación, la cantidad de tiempo del seguimiento no se relaciona con la probabilidad de desarrollar el resultado. Éste será el caso si el seguimiento concluye en una fecha fija o a una edad concreta. La pérdida durante el seguimiento ocurre cuando los participantes abandonan el estudio antes de esa fecha, lo cual puede afectar la validez del estudio si la pérdida del seguimiento ocurre selectivamente en

los individuos expuestos o en personas con alto riesgo de desarrollar la enfermedad (“censura informativa”). En el ejemplo anterior, los pacientes perdidos durante el seguimiento en programas de tratamiento sin seguimiento activo tenían menos células CD4 colaboradoras que aquellos que permanecieron bajo observación y que por consiguiente presentaban un riesgo más alto de morir¹¹⁶.

Es importante distinguir a las personas que finalizan el estudio de aquellas que se pierden durante el seguimiento. Desafortunadamente, el programa estadístico no suele distinguir entre las dos situaciones: en ambos casos el tiempo de seguimiento es suspendido automáticamente (“censurado”) al final del periodo de observación. Por lo tanto, los investigadores necesitan decidir, idealmente en la fase de planificación del estudio, cómo manejarán las pérdidas durante el seguimiento. Cuando se pierden pocos pacientes, los investigadores pueden excluir a los individuos con seguimiento incompleto o tratarlos como si hubieran abandonado con vida en el momento de la pérdida o al final del estudio. Recomendamos a los autores que indiquen cuántos pacientes se perdieron durante el seguimiento y qué estrategias de censura utilizaron.

12 (d). *Estudios de casos y controles: si procede, explique cómo se parearon casos y controles*

Ejemplo

“Empleamos la prueba McNemar, la prueba *t* para muestras pareadas y el análisis de regresión logística condicional para comparar la ocurrencia de embolias cerebrales espontáneas, enfermedad carotídea y desviación de la circulación venosa a la arterial en los pacientes con demencia y en sus controles pareados por factores de riesgo cardiovascular”¹¹⁷.

Explicación

En los estudios de casos y controles pareados individualmente, el análisis crudo de la *odds ratio*, ignorando el emparejamiento, suele sesgar la estimación hacia la unidad (véase el **Recuadro 2**). Por lo tanto, es necesario un análisis pareado. Intuitivamente esto puede entenderse como un análisis estratificado: cada caso se ve como un estrato con su conjunto de controles pareados. El análisis se basa en considerar si el caso está expuesto más a menudo que los controles, a pesar de ser muy parecidos respecto a las variables de pareamiento. Los investigadores pueden hacer tal análisis estratificado mediante el método de Mantel-Haenszel en una tabla “pareada” de 2 por 2. En su forma más simple, la *odds ratio* se convierte en la razón de los pares discordantes para la variable de exposición. Si el emparejamiento se hizo por variables como la edad y el sexo, que son atributos universales, el análisis no requiere retener el pareamiento individual, persona a persona, y un simple análisis por categorías de edad y sexo es suficiente⁵⁰. Sin embargo, para otras variables de pareamiento, como en el vecindario, la hermandad o la amistad, cada conjunto pareado debe ser considerado su propio estrato.

En los estudios pareados individualmente, el método de análisis más usado es la regresión logística condicional, en la cual se consideran juntos cada caso y sus controles. El método condicional es necesario cuando el número de controles varía entre los casos, y cuando se requiere ajustar el modelo por otras variables además de las de pareamiento. Para permitir a los lectores juzgar si en el análisis se tuvo en cuenta apropiadamente el diseño pareado, recomendamos que los autores describan con detalle los métodos estadísticos que emplearon para analizar los datos. Si al tener en cuenta el pareamiento tiene un efecto pequeño en las estimaciones, los autores pueden optar por presentar un análisis no pareado.

12 (d). Estudios transversales: si procede, especifique cómo se tiene en cuenta en el análisis la estrategia de muestreo

Ejemplo

“Se calcularon los errores estándar (EE) usando el método de expansión de Taylor para estimar los errores de muestreo de los estimadores, basados en el diseño de muestra complejo. (...) Se encontró que el efecto de diseño general para la presión sanguínea diastólica fue de 1,9 para los hombres y 1,8 para las mujeres, y para la presión sistólica fue de 1,9 para los hombres y 2,0 para las mujeres¹¹⁸.”

Explicación

La mayoría de los estudios transversales utilizan una estrategia de muestreo preespecificada para seleccionar a los participantes de una población fuente. El muestreo puede ser más complejo que meramente tomar una muestra aleatoria simple; puede incluir varias etapas y grupos de participantes (por ejemplo, en distritos o pueblos). La estratificación proporcional puede asegurar que los subgrupos con una característica específica estén representados correctamente. La estratificación desproporcionada puede ser útil para sobremuestrear un subgrupo de particular interés.

Una asociación estimada basada en una muestra compleja puede ser más o menos precisa que la derivada de una muestra aleatoria simple. Se deben corregir las medidas de precisión, como el error estándar o el intervalo de confianza, usando el efecto de diseño, una medida de razón que describe cuánta precisión se gana o se pierde si se usa una estrategia de muestreo más compleja en lugar de un muestreo aleatorio simple¹¹⁹. La mayoría de las técnicas de muestreo complejas disminuyen la precisión, resultando en un efecto de diseño mayor de 1.

Aconsejamos que los autores expongan claramente el método empleado para ajustar por las estrategias de muestreo complejas, de modo que los lectores puedan entender cómo el método de muestreo elegido influyó en la precisión de las estimaciones obtenidas. Por ejemplo, con muestras conglomeradas, el intercambio implícito entre una recogida de datos más fácil y una pérdida de precisión es transparente si se indica el efecto de diseño. En el ejemplo, el efecto de diseño calculado, de 1,9 para los hombres, indica que el tamaño de la muestra real necesitaría ser 1,9 veces mayor que con el muestreo aleatorio simple para que las estimaciones resultantes fueran igualmente precisas.

12 (e). Describa los análisis de sensibilidad

Ejemplo

“Debido a que tuvimos una proporción relativamente alta de pacientes muertas «perdidas» con datos insuficientes (38/148 = 25,7%) en comparación con las pacientes vivas (15/437 = 3,4%) (...), es posible que esto sesgara los resultados. Por consiguiente, se llevó a cabo un análisis de sensibilidad. Asumimos que la proporción de mujeres que usan anticonceptivos orales en el grupo del estudio se aplica a todas (19,1% para muertas y 11,4% para pacientes vivas), y entonces se consideraron dos posibilidades extremas: que todas las pacientes perdidas expuestas usaron píldoras de segunda generación o que todas usaron píldoras de tercera generación”¹²⁰.

Explicación

Los análisis de sensibilidad son útiles para investigar si los resultados principales son consistentes o no con los obtenidos con estrategias o suposiciones de análisis alternativas¹²¹. Los problemas que pueden examinarse incluyen los criterios para la inclusión en los análisis, las definiciones de exposiciones o resultados¹²², qué variables de confusión requieren ajustarse, el

Recuadro 6–Datos ausentes: problemas y posibles soluciones.

Un planteamiento común para tratar los datos ausentes es restringir el análisis a los individuos que tengan los datos completos de todas las variables requeridas para un análisis en particular. Si bien en muchas circunstancias los análisis con “casos completos” no están sesgados, también pueden estarlo y siempre son ineficientes¹⁰⁸. El sesgo surge si los individuos con datos ausentes no son representativos de toda la muestra. La ineficiencia puede ocurrir debido a una reducción en el tamaño de muestra que será analizado.

Emplear la última observación en un diseño de medidas repetidas puede distorsionar las tendencias en el tiempo si las personas que intuyen el evento resultante abandonan selectivamente el estudio¹⁰⁹. Insertar un indicador de dato ausente para un confusor puede aumentar la confusión residual¹⁰⁷. La imputación, en la cual cada valor ausente es remplazado por un valor asumido o estimado, puede conducir a una atenuación o una exageración de la asociación de interés, y sin el uso de métodos sofisticados, como los que se describen más adelante, pueden producirse errores estándar muy pequeños.

Rubin^{108,110} desarrolló una tipología de problemas con datos ausentes basándose en un modelo para la probabilidad de que se pierda una observación. Los datos se describen como ausentes completamente al azar si la probabilidad de que una observación particular esté ausente no depende del valor de ninguna otra variable observable. Los datos son ausentes al azar si, considerando los datos observados, la probabilidad de que la observación esté ausente es independiente de los valores verdaderos del dato ausente. Por ejemplo, supongamos que los niños pequeños son más propensos a medidas ausentes de espirometría, pero que la probabilidad de ausencia no se relaciona con la función pulmonar real no observada, después de ajustar por la edad. Así, la medida de la función pulmonar sería del tipo ausentes al azar en los modelos que incluyen la edad. Los datos son ausentes no aleatoriamente si la probabilidad de ausencia depende del valor ausente aun después de tener en cuenta los datos disponibles. Cuando los datos son del tipo ausentes no aleatoriamente, para obtener inferencias válidas se requieren suposiciones explícitas sobre los mecanismos por los cuales se perdieron los datos.

Los métodos para tratar con datos ausentes al azar pueden ser de tres clases^{108,111}: aproximaciones basadas en verosimilitud¹¹², estimación ponderada¹¹³ e imputación múltiple^{111,114}. De las tres, la imputación múltiple es la más usada y flexible, en especial cuando diversas variables tienen valores ausentes¹¹⁵. Los resultados que se obtienen al utilizar cualquiera de estas tres opciones deben compararse con los derivados de análisis de casos completos, y comentar las diferencias importantes. La plausibilidad de las suposiciones realizadas en análisis con datos ausentes generalmente no puede verificarse. En particular, es imposible probar que los datos son ausentes al azar más que ausentes no aleatoriamente. Por tanto, tales análisis son mejor apreciados con base en el análisis de sensibilidad (véanse los puntos 12e y 17).

manejo de datos ausentes^{120,123}, el posible sesgo de selección o los sesgos derivados de la medida imprecisa o inconsistente de la exposición, la enfermedad y otras variables, así como decisiones de análisis específicas, como el tratamiento de las variables cuantitativas (ver el punto 11). Es creciente el empleo de métodos sofisticados para modelar simultáneamente la influencia de varios sesgos o suposiciones^{124–126}.

En 1959, Cornfield et al mostraron que un riesgo relativo de 9 para fumadores de cigarrillos y cáncer de pulmón era sumamente

improbable que se debiera a cualquier confusor concebible, debido a que el confusor necesitaría prevalecer por lo menos nueve veces más en los fumadores que en los no fumadores¹²⁷. Este análisis no excluye la posibilidad de que un factor estuviera presente, pero identifica el predominio que tal factor necesitaría tener. La misma aproximación se utilizó para identificar probables factores confusores que podrían explicar la asociación entre la leucemia en la niñez y el vivir cerca de líneas de fuerza eléctrica¹²⁸. Generalmente los análisis de sensibilidad pueden usarse para identificar el grado de confusión, el sesgo de selección o el sesgo de información requeridos para distorsionar una asociación. Poco conocido, un uso importante del análisis de sensibilidad es cuando un estudio muestra una asociación pequeña o nula entre una exposición y un resultado, y es plausible que estén ocurriendo confusión o sesgos hacia el valor nulo.

Resultados

La sección de resultados debe hacer un claro recuento de lo que se encontró, del reclutamiento de los participantes, de la descripción de la población del estudio y de los principales resultados del análisis. Ha de estar libre de interpretaciones e ideas que reflejen las opiniones y los puntos de vista de los autores.

13. Participantes

13 (a). Indique el número de participantes en cada fase del estudio; p. ej., número de participantes elegibles, analizados para ser incluidos, confirmados elegibles, incluidos en el estudio, los que tuvieron un seguimiento completo y los analizados

Ejemplo

“De la muestra de 105 bares y tabernas, 13 establecimientos ya no funcionaban como tales y 9 se localizaban en restaurantes, quedando entonces 83 negocios elegibles. En 22 casos, el propietario no pudo ser localizado por teléfono después de por lo menos seis intentos. Los propietarios de 36 bares rehusaron participar en el estudio. (...) Los 25 bares y tabernas que participaron empleaban a 124 camareros, 67 de ellos trabajaban por lo menos 24 horas a la semana. Cincuenta y cuatro de los camareros con horario diurno (81%) completaron la entrevista y la espirometría basal; 53 de ellos (98%) terminaron el seguimiento”¹²⁹.

Explicación

La información detallada sobre el proceso de reclutamiento de los participantes del estudio es importante por varias razones. Los incluidos en un estudio a menudo difieren de la población diana a la cual se aplican los resultados. Esto puede resultar en estimaciones de la prevalencia o la incidencia que no reflejan lo que ocurre en la base poblacional. Por ejemplo, la gente que aceptó participar en una encuesta por correo sobre el comportamiento sexual asistía a la iglesia con menor frecuencia, tenía actitudes sexuales menos conservadoras y menor edad en su primera relación sexual, y era más probable que fumara y bebiera alcohol que la gente que rehusó participar¹³⁰. Estas diferencias sugieren que las encuestas por correo pueden sobrestimar actitudes liberales y actividad sexual en la población. Tal sesgo de respuesta (véase el Recuadro 3) puede distorsionar asociaciones exposición-enfermedad si las asociaciones difieren entre los elegibles para el estudio y los incluidos. Otro ejemplo es la asociación encontrada entre una corta edad materna y el desarrollo de leucemia en los descendientes, que se ha observado en algunos estudios de casos y controles^{131,132} y ha sido explicada

por la participación diferenciada de mujeres jóvenes en el grupo de casos y el grupo de controles. Las mujeres jóvenes con niños sanos era menos probable que participaran que las madres jóvenes con niños enfermos¹³³. Si bien la baja participación no afecta necesariamente la validez de un estudio, la información clara sobre la participación y las razones de la no participación son esenciales. Del mismo modo, como no hay definiciones globalmente aceptadas sobre participación, respuesta o tasa de seguimiento, los lectores necesitan entender cómo los autores calculan tales proporciones¹³⁴.

Idealmente los investigadores deben especificar el número de individuos considerados en cada fase del reclutamiento de los participantes del estudio y la selección de la base poblacional para la inclusión de los datos de los participantes en el análisis. Dependiendo del tipo de estudio, se puede incluir el número de individuos considerados elegibles, el número determinado por elegibilidad, el número incluido en el estudio, el número examinado, el número seguido y el número incluido en el análisis. Si el muestreo de los participantes del estudio se realiza en dos o más etapas, como en el ejemplo de arriba (muestreo multietápico), se puede requerir la información sobre diversas unidades de muestreo. En estudios de casos y controles, aconsejamos que los autores indiquen el flujo de participantes por separado para el grupo de casos y el de controles¹³⁵. A veces los controles se pueden seleccionar de varias fuentes, incluyendo por ejemplo los pacientes hospitalizados y los habitantes de la comunidad. En tal caso recomendamos un recuento separado del número de participantes para cada tipo de grupo control. Olson et al. propusieron directrices útiles para informar sobre los controles reclutados mediante selección aleatoria por número de teléfono y otros métodos¹³⁶.

Un trabajo reciente¹³⁷ sobre estudios epidemiológicos publicados en 10 revistas de epidemiología general, salud pública y medicina, encontró que la información respecto a la participación fue proporcionada en 47 de 107 estudios de casos y controles (59%), 49 de 154 estudios de cohortes (32%) y 51 de 86 estudios transversales (59%). La información incompleta o ausente respecto a la participación o la no participación en estudios epidemiológicos también ha sido documentada en otros dos artículos^{4,5}. Finalmente, hay pruebas de que la participación en estudios epidemiológicos puede haber disminuido en las últimas décadas^{137,138}, lo que subraya la necesidad de informes transparentes¹³⁹.

13 (b). Describa las razones de la pérdida de participantes en cada fase

Ejemplo

“Las principales razones de la no participación fueron que las participantes estaban demasiado enfermas o habían muerto antes de la entrevista (30% de los casos, <1% de los controles), sin respuesta (2% de los casos, 21% de los controles), rehusaron (10% de los casos, 29% de los controles) y otras razones (negativa de participar del médico especialista o médico de familia, no hablar inglés, problemas mentales) (7% de los casos, 5% de los controles)”¹⁴⁰.

Explicación

La explicación de las razones por las que la mayoría de la gente no participó en un estudio, o de por qué fueron excluidos del análisis estadístico, ayuda a los lectores a juzgar si la población estudiada era representativa de la base poblacional y si posiblemente se introdujo un sesgo. Por ejemplo, en una encuesta transversal de salud, la no participación debida a razones no relacionadas con el estado de salud (por ejemplo, la carta de invitación no fue entregada porque la dirección era incorrecta)

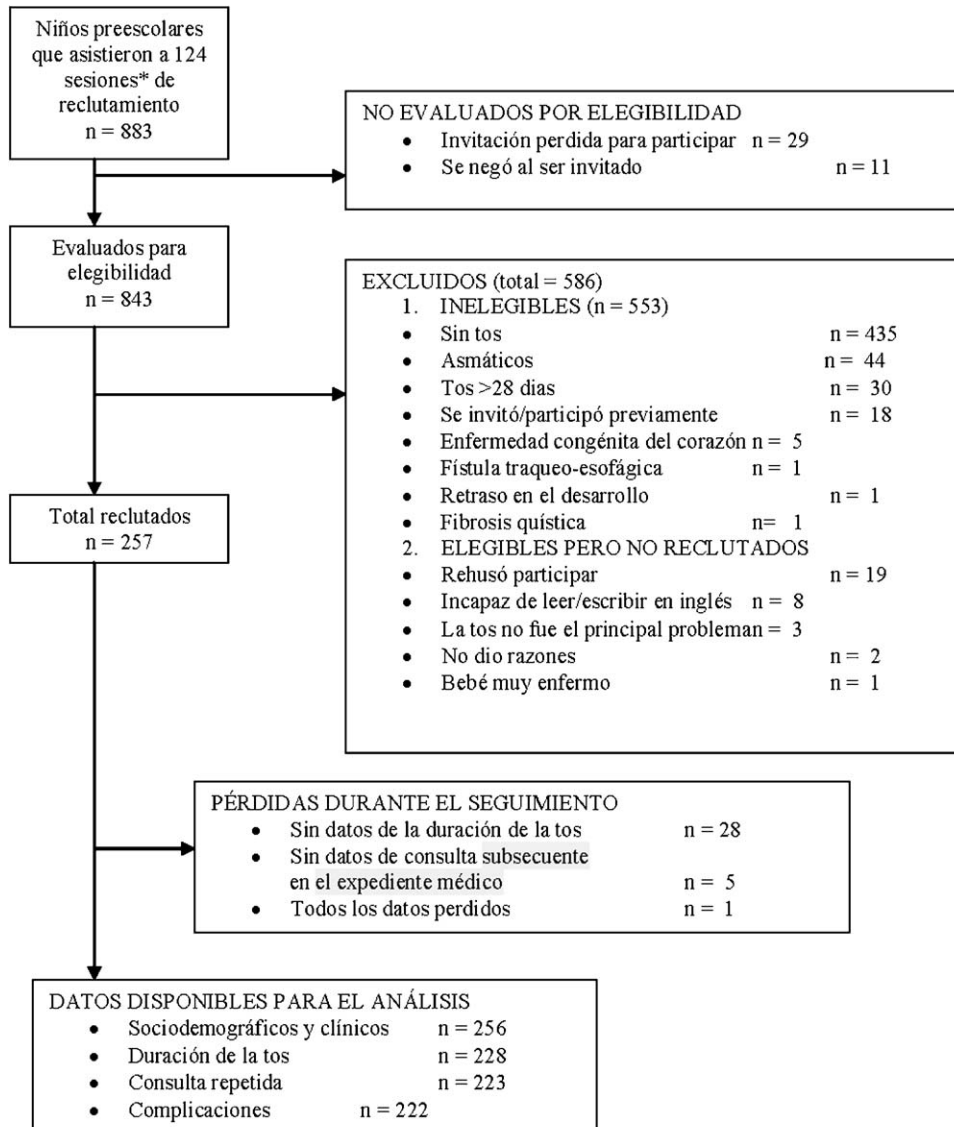
afectará la precisión de las estimaciones, pero probablemente no introducirá un sesgo. Por el contrario, si muchos individuos optan por no participar debido a enfermedad o a percepción de buena salud, los resultados pueden subestimar o sobrestimar la prevalencia de la mala salud en la población.

13 (c). Considere el uso de un diagrama de flujo

Ejemplo

Explicación

Un diagrama de flujo informativo y bien estructurado puede mostrar fácil y claramente la información que, de otra manera, puede requerir una descripción muy larga¹⁴², como en el ejemplo anterior. Un diagrama útil puede incluir los principales resultados, tales como el número de acontecimientos para el resultado primario. Recomendamos el uso de diagramas de flujo, particularmente para los estudios observacionales complejos, pero no proponemos un formato específico para ellos.



*Datos perdidos del denominador de una sesión en cuál al menos 3 se atendieron con tos, 2 reclutados
Diagrama de flujo de Hay et al.¹⁴¹.

14. Datos descriptivos

14 (a). *Describe las características de los participantes en el estudio (p. ej., demográficas, clínicas, sociales) y la información sobre las exposiciones y los posibles factores de confusión*

Ejemplo

Tabla adaptada de Osella et al¹⁴³.

Características de la base poblacional en el momento del reclutamiento. Castellana G (Italia), 1985–1986

	VHC negativo n = 1458	VHC positivo n = 511	Desconocido n = 513
Sexo (%)			
Hombres	936 (64%)	296 (58%)	197 (39%)
Mujeres	522 (36%)	215 (42%)	306 (61%)
Media de la edad al inicio (DE)	45,7(10,5)	52,0 (9,7)	52,5 (9,8)
Consumo diario de alcohol (%)			
Nada	250 (17%)	129 (25%)	119 (24%)
Moderado ^a	853 (59%)	272 (53%)	293 (58%)
Excesivo ^b	355 (24%)	110 (22%)	91 (18%)

VHC: virus de la hepatitis C.

^a Hombres < 60 g de etanol /día, mujeres < 30 g de etanol/día.

^b Hombres > 60 g de etanol /día, mujeres > 30 g de etanol/día.

Explicación

Los lectores necesitan la descripción de los participantes del estudio y de sus exposiciones para juzgar la posible generalización de los resultados. La información sobre los posibles confusores, incluyendo si se consideraron y cómo se determinaron, influye en los juicios sobre la validez del estudio. Aconsejamos a los autores resumir las variables continuas para cada grupo de estudio mediante la media y la desviación estándar, o si los datos tienen una distribución asimétrica, como sucede a menudo, la mediana y el rango en percentiles (por ejemplo, percentiles 25 y 75). Las variables nominales formadas por una pequeña cantidad de categorías (como las fases de la enfermedad I a IV) no se deben presentar como variables continuas; es preferible dar cantidades y proporciones para cada categoría (véase también el Recuadro 4). En estudios que comparan grupos, las características y los datos descriptivos se deben presentar por grupo, como en el ejemplo anterior.

Las medidas inferenciales, como los errores estándar y los intervalos de confianza, no se deben utilizar para describir la variabilidad de las características, y en las tablas descriptivas se deben evitar las pruebas de significación. Tampoco los valores *p* son un criterio apropiado para seleccionar por qué confusores ajustar el análisis; incluso pequeñas diferencias en un confusor que tenga un fuerte efecto sobre el resultado pueden ser importantes^{144,145}.

En estudios de cohortes puede ser útil documentar cómo se relaciona una exposición con otras características y con los posibles confusores. Los autores podrían presentar esta información en una tabla con columnas para los participantes en dos o más categorías de exposición, que permita evaluar las diferencias con respecto a los confusores entre estas categorías.

En estudios de casos y controles, los posibles confusores no se pueden apreciar mediante la comparación de los casos y los controles. Los controles representan la población fuente y normalmente serán diferentes a los casos en muchos aspectos. Por ejemplo, en un estudio sobre anticonceptivos orales e infarto de miocardio, una muestra de mujeres jóvenes con infarto

presentará más factores de riesgo para esa enfermedad que el grupo control, tales como colesterol sérico alto, tabaquismo y antecedentes familiares positivos¹⁴⁶. Esto no influye en la evaluación del efecto de los anticonceptivos orales en tanto que su prescripción no estuvo guiada por la presencia de los factores de riesgo, por ejemplo debido a que estos factores de riesgo se establecieron después de ocurrido el evento (véase también el Recuadro 5). En los estudios de casos y controles, el equivalente de comparar la presencia de posibles confusores (como se hace en las cohortes) entre expuestos y no expuestos se puede realizar explorando la población fuente de los casos: si el grupo control es suficientemente grande y representa a la población fuente, los controles expuestos y no expuestos se pueden comparar con respecto a los posibles confusores^{121,147}.

14 (b). *Indique el número de participantes con datos ausentes en cada variable de interés*

Ejemplo

Tabla adaptada de Hay et al¹⁴¹.

Puntos sintomáticos terminales empleados en el análisis de supervivencia			
	Tos	Disnea	Somnolencia
Síntoma resuelto	201 (79%)	138 (54%)	171 (67%)
Censurado	27 (10%)	21 (8%)	24 (9%)
Nunca sintomático	0	46 (18%)	11 (4%)
Datos ausentes	28 (11%)	51 (20%)	50 (20%)
Total	256 (100%)	256 (100%)	256 (100%)

Explicación

Puesto que los datos ausentes pueden sesgar o afectar la generalización de los resultados, los autores deben informar a los lectores sobre la cantidad de datos perdidos para las exposiciones, los posibles confusores y otras características importantes de los pacientes (véanse también el punto 12c y el Recuadro 6). En un estudio de cohortes, los autores deben especificar la cantidad de pérdidas durante el seguimiento (con sus razones), pues un seguimiento incompleto puede sesgar los hallazgos (véanse también los puntos 12d y 13)¹⁴⁸. Recomendamos a los autores usar tablas y figuras para enumerar los datos ausentes.

14 (c). *Estudios de cohortes: resume el periodo de seguimiento (p. ej., promedio y total)*

Ejemplo

“Durante el seguimiento de 4366 personas-año (mediana 5,4, máximo 8,3 años), 265 sujetos fueron diagnosticados como dementes, incluyendo 202 con enfermedad de Alzheimer”¹⁴⁹.

Explicación

Los lectores deben saber la duración y la extensión del seguimiento de los datos disponibles acerca del evento resultado. Los autores pueden presentar un resumen del promedio de seguimiento incluyendo la media, la mediana o ambas. La media permite al lector calcular el número total de personas-año multiplicándola por el número de participantes en el estudio. También se deben presentar los tiempos mínimo y máximo, o los percentiles de la distribución, para mostrar a los lectores la dispersión del tiempo de seguimiento. Deben especificar el total personas-año de seguimiento o alguna indicación de la proporción de datos que fueron capturados con respecto al total posible¹⁴⁸. Toda esta información sobre los participantes se debe presentar

separadamente en dos o más categorías de exposición. Casi la mitad de 132 artículos publicados en revistas sobre cáncer (en su mayoría estudios de cohortes) no presentaban algún resumen de la duración del seguimiento³⁷.

15. Datos de las variables de resultado

Estudios de cohortes: indique el número de eventos resultado o bien proporcione medidas de resumen a lo largo del tiempo

Ejemplo

Tabla adaptada de Kengeya-Kayondo et al.¹⁵⁰

Tasas de seroconversión VIH-1 por variables sociodemográficas seleccionadas: 1990–1993

Variable	Personas-año	N.º seroconvertidos	Tasa/1000 personas-año (IC95%)
Año			
1990	2197,5	18	8,2 (4,4–12,0)
1991	3210,7	22	6,9 (4,0–9,7)
1992	3162,6	18	5,7 (3,1–8,3)
1993	2912,9	26	8,9 (5,5–12,4)
1994	1104,5	5	4,5 (0,6–8,5)
Tribu			
Bugandesa	8433,1	48	5,7 (4,1–7,3)
Otras ugandesas	578,4	9	15,6 (5,4–25,7)
Ruandesa	2318,6	16	6,9 (3,5–10,3)
Otra tribu	866,0	12	13,9 (6,0–21,7)
Religión			
Musulmana	3313,5	9	2,7 (0,9–4,5)
Otra	8882,7	76	8,6 (6,6–10,5)

IC95%: intervalo de confianza del 95%.

Estudios de casos y controles: indique el número de participantes en cada categoría de exposición o bien proporcione medidas resumen de exposición

Ejemplo

Tabla adaptada de Mastrangelo et al.¹⁵¹

Exposición entre casos de cirrosis hepática y controles		
	Casos (n = 40)	Controles (n = 139)
Cloruro de vinilo monómero (exposición acumulada: ppm × años)		
< 160	7 (18%)	38 (27%)
160–500	7 (18%)	40 (29%)
500–2500	9 (23%)	37 (27%)
> 2500	17 (43%)	24 (17%)
Consumo de alcohol (g/día)		
< 30	1 (3%)	82 (59%)
30–60	7 (18%)	46 (33%)
> 60	32 (80%)	11 (8%)
AgsHB/VHC		
Negativo	33 (83%)	136 (98%)
Positivo	7 (18%)	3 (2%)

AgsHB: antígeno de superficie de la hepatitis B; VHC: virus de la hepatitis C.

Estudios transversales: indique el número de eventos resultado o bien proporcione medidas resumen.

Ejemplo

Tabla adaptada de Salo et al.¹⁵².

Prevalencia de asma actual y de fiebre del heno diagnosticadas por nivel promedio de antígeno de <i>Alternaria alternata</i> en el hogar				
Nivel categorizado de <i>Alternaria</i> ^a	Asma actual		Fiebre del heno diagnosticada	
	n	Prevalencia ^b (IC95%)	n	Prevalencia ^b (IC95%)
1 ^{er} tercil	40	4,8 (3,3–6,9)	93	16,4 (13,0–20,5)
2 ^o tercil	61	7,5 (5,2–10,6)	122	17,1 (12,8–22,5)
3 ^{er} tercil	73	8,7 (6,7–11,3)	93	15,2 (12,1–18,9)

^a 1^{er} tercil < 3,90 µg/g; 2^o tercil 3,90–6,27 µg/g; 3^{er} tercil ≥ 6,28 µg/g.

^b Porcentaje (IC95%) ponderado por el diseño de muestreo polifásico de la Encuesta Nacional de Plomo y Alergenos en el Hogar.

Explicación

Antes de establecer la posible asociación entre las exposiciones (factores de riesgo) y los resultados, los autores deben proporcionar los datos descriptivos relevantes. Puede ser posible y significativo presentar medidas de asociación en la misma tabla que muestra los datos descriptivos (véase el punto 14a). En un estudio de cohortes con eventos como resultados, indique el número de eventos para cada resultado de interés. Considere especificar la tasa del evento por personas-año de seguimiento. Si el riesgo de un evento cambia durante el tiempo de seguimiento, presente las cifras y las tasas de eventos en intervalos apropiados del seguimiento, como una tabla de vida de Kaplan-Meier o de manera gráfica. Puede ser preferible realizar gráficas de incidencia acumulada que partan de 0% en lugar de que disminuyan a partir del 100%, especialmente si la tasa del evento es menor, por ejemplo, del 30%¹⁵³. Considere presentar esta información por separado para los participantes en diversas categorías de la exposición de interés. Si un estudio de cohortes está investigando otros resultados relacionados con el tiempo (p.ej., marcadores cuantitativos de una enfermedad, como la presión sanguínea), presente medidas resumen apropiadas (por ejemplo, medias y desviaciones estándar) a lo largo del tiempo, quizás en una tabla o una figura.

Para los estudios transversales recomendamos presentar el mismo tipo de información sobre eventos resultado prevalentes o medidas de resumen. Para los estudios de casos y controles se indicarán las exposiciones para los casos y los controles por separado, como frecuencias o medidas resumen cuantitativas¹⁵⁴. Para todos los diseños, también puede resultar útil tabular los eventos resultado o las exposiciones continuas en categorías, incluso si los datos no se analizan como tales.

16. Resultados principales

16 (a). *Proporcione estimaciones no ajustadas y, si procede, ajustadas por factores de confusión, así como su precisión (p. ej., intervalos de confianza del 95%). Especifique los factores de confusión por los que se ajusta y las razones para incluirlos*

Ejemplo 1

“Inicialmente consideramos las siguientes variables como posibles confusores mediante un análisis estratificado de Mantel-Haenszel: (...) las variables que incluimos en los modelos de regresión logística finales fueron aquellas (...) que producían un cambio del 10% en la *odds ratio* después del ajuste de Mantel-Haenszel”¹⁵⁵.

Ejemplo 2

Tabla adaptada de Tilhonen et al¹⁵⁶.

Tasas relativas de rehospitalización por tratamiento en pacientes de atención primaria después de la primera hospitalización por esquizofrenia o un trastorno esquizoafectivo

Tratamiento	N.º de recaídas	Personas-año	Tasa relativa bruta (IC95%)	Tasa relativa ajustada (IC95%)	Tasa relativa completamente ajustada (IC95%)
Perfenazina	53	187	0,41 (0,29–0,59)	0,45 (0,32–0,65)	0,32 (0,22–0,49)
Olanzapina	329	822	0,59 (0,45–0,75)	0,55 (0,43–0,72)	0,54 (0,41–0,71)
Clozapina	336	804	0,61 (0,47–0,79)	0,53 (0,41–0,69)	0,64 (0,48–0,85)
Clorprotixeno	79	146	0,79 (0,58–1,09)	0,83 (0,61–1,15)	0,64 (0,45–0,91)
Tioridazina	115	201	0,84 (0,63–1,12)	0,82 (0,61–1,10)	0,70 (0,51–0,96)
Perfenazina	155	327	0,69 (0,58–0,82)	0,78 (0,59–1,03)	0,85 (0,63–1,13)
Risperidona	343	651	0,77 (0,60–0,99)	0,80 (0,62–1,03)	0,89 (0,69–1,16)
Haloperidol	73	107	1,00	1,00	1,00
Clorpromazina	82	127	0,94 (0,69–1,29)	0,97 (0,71–1,33)	1,06 (0,76–1,47)
Levomepromazina	52	63	1,21 (0,84–1,73)	0,82 (0,58–1,18)	1,09 (0,76–1,57)
Fármacos no antipsicóticos	2.248	3.362	0,98 (0,77–1,23)	1,01 (0,80–1,27)	1,16 (0,91–1,47)

Ajustado por sexo, año, edad al inicio del seguimiento, número de recaídas previas, duración de la primera hospitalización y duración del seguimiento (columna ajustada), y adicionalmente por una calificación de la propensión a iniciar un tratamiento diferente al haloperidol (columna totalmente ajustada).

Explicación

En muchas situaciones los autores pueden presentar los resultados del análisis sin ajustar o mínimamente ajustados, y análisis completamente ajustados. Recomendamos presentar el análisis no ajustados junto con los datos principales, por ejemplo el número de casos y de controles expuestos y no expuestos. Esto permite que el lector comprenda los datos que están detrás de las medidas de la asociación (véase también el punto 15). Para los análisis ajustados, indique el número de personas en el análisis, ya que este número puede diferir debido a valores ausentes en las covariables (véase también el punto 12c). Los estimadores se deben proporcionar con intervalos de confianza.

Los lectores pueden comparar las medidas de asociación no ajustadas con las ajustadas por posibles confusores y juzgar cuánto se modifican y en qué dirección. Pueden pensar que los resultados “ajustados” equivalen a la parte causal de la medida de asociación, pero estos resultados no necesariamente están libres de un error de muestreo aleatorio, un sesgo de selección o una confusión residual (véase el Recuadro 5). Así, se debe tener mucho cuidado cuando se interpreten los resultados ajustados, ya que su validez a menudo depende crucialmente del conocimiento completo de los confusores importantes, de su determinación precisa y de la especificación apropiada en el modelo estadístico (véase también el punto 20)^{157,158}.

Los autores deben explicar todos los posibles confusores considerados y los criterios para excluir o incluir variables en los modelos estadísticos. Las decisiones acerca de la exclusión o la inclusión de variables se deben guiar por el conocimiento o por suposiciones explícitas sobre las relaciones causales. Las decisiones incorrectas pueden introducir sesgos, por ejemplo incluyendo las variables que están en la línea causal entre la exposición y la enfermedad (a menos que el propósito sea evaluar qué parte del efecto está producida por la variable intermediaria). Si la decisión para incluir una variable en el modelo se basó en un cambio en el estimador, es importante indicar qué cambio se consideró suficientemente importante para justificar su inclusión. Si se empleó una estrategia de modelado mediante “eliminación hacia atrás” o “inclusión hacia adelante” para la selección de los confusores, explique el proceso e indique el nivel de significación para rechazar la hipótesis nula en ausencia de confusores. Cabe destacar que nosotros, al igual que otros autores, no recomendamos seleccionar los confusores basándose únicamente en la significación estadística^{147,159,160}.

Estudios recientes sobre la calidad de los artículos de estudios epidemiológicos han hallado que la mayoría especifican los intervalos de confianza⁴. Sin embargo, pocos autores explican la selección de sus variables confusoras^{4,5}.

16 (b). Si categoriza variables continuas, describa los límites de los intervalos

Ejemplo

Tabla adaptada de Sagiv et al¹⁶¹.

Bifenilos policlorados en suero de cordón umbilical		
Cuartil	Rango (ng/g)	Cantidad
1	0,07–0,24	180
2	0,24–0,38	181
3	0,38–0,60	181
4	0,61–18,14	180

Explicación

Categorizar datos continuos tiene varias implicaciones importantes para el análisis (véase el Recuadro 4) y también afecta la presentación de los resultados. Empleando tablas, los resultados se deben mostrar para cada categoría de la exposición, por ejemplo como número de personas en riesgo, personas-tiempo en riesgo, y si es relevantes por separado para cada grupo (p.ej., los casos y los controles). Los detalles de las categorías usadas pueden ayudar a la comparación de estudios y a los metaanálisis. Si los datos se han agrupado usando puntos de corte convencionales, como los límites del índice de masa corporal¹⁶², los límites grupales (p.ej., los intervalos de valores) se pueden derivar fácilmente, con excepción de las categorías más altas y más bajas. Si se utilizan categorías derivadas de percentiles, los límites de la categoría no se pueden deducir de los datos. Como mínimo, los autores deben indicar los límites de la categoría; también es útil especificar el intervalo de los datos y la media o la mediana dentro de las categorías.

16 (c). Si fuera pertinente, valore acompañar las estimaciones del riesgo relativo con estimaciones del riesgo absoluto para un periodo de tiempo relevante

Ejemplo

“Se estimó que el uso de terapia hormonal sustitutiva durante 10 años resulta en cinco (IC95%: 3–7) cánceres de mama adicionales por cada 1000 usuarias de preparaciones de estrógenos solos y en 19 (15–23) cánceres adicionales por 1000 usuarias de combinaciones de estrógenos-progestágenos”¹⁶³.

Explicación

Los resultados de los estudios que investigan la asociación entre una exposición y una enfermedad se presentan comúnmente en términos relativos, como razones de riesgos, de tasas o de odds (véase el Recuadro 8). Las medidas relativas indican la fuerza de la asociación entre una exposición y una enfermedad. Si el riesgo relativo está lejos de 1, es menos probable que la asociación se deba a un factor de confusión^{164,165}. Los efectos relativos o las asociaciones tienden a ser más consistentes a lo largo de estudios y de poblaciones que las medidas absolutas, pero lo que a menudo tiende a ser el caso puede ser irrelevante en una situación particular. Por ejemplo, se obtuvieron riesgos relativos similares para los factores de riesgo cardiovasculares clásicos para los hombres que vivían en Irlanda del Norte, Francia, Estados Unidos y Alemania, a pesar de que el riesgo subyacente de enfermedad

cardíaca coronaria varía sustancialmente entre estos países^{166,167}. En cambio, en un estudio sobre la hipertensión como factor de riesgo para la mortalidad por enfermedad cardiovascular, los datos eran más compatibles con una diferencia de tasas constante que con una razón de tasas constante¹⁶⁸.

Los modelos estadísticos que más se utilizan, incluyendo las regresiones logística¹⁶⁹ y proporcional de riesgos (Cox)¹⁷⁰, se basan en medidas de razón. En estos modelos se distinguen fácilmente las desviaciones de las medidas de efecto con respecto a la constante. Sin embargo, las medidas que evalúan desviaciones de la aditividad con respecto a las diferencias de riesgo, como el exceso de riesgo relativo de la interacción (RERI, véanse el punto 12b y el Recuadro 8), se pueden estimar a partir de los modelos basados en medidas de razón.

Recuadro 7–Medidas de asociación, efecto e impacto.

Los estudios observacionales sólo se pueden llevar a cabo para describir la magnitud y la distribución de un problema de salud en la población. Pueden investigar el número de personas que tienen una enfermedad en un momento particular (prevalencia) o que la desarrollan en un periodo de tiempo definido (incidencia). La incidencia se puede expresar como la proporción de gente que desarrolla la enfermedad (incidencia acumulada) o como la tasa por personas-tiempo de seguimiento (tasa de incidencia). Se emplean términos específicos para describir diversas incidencias; entre otros, tasa de mortalidad, tasa de natalidad, tasa de ataque y tasa de fatalidad de casos. Del mismo modo, los términos como prevalencia puntual y de periodo, prevalencia anual o del curso de la vida, se utilizan para describir diversos tipos de prevalencia³⁰.

Otros estudios observacionales tratan relaciones de causa-efecto. Su objetivo es comparar el riesgo, la tasa o la prevalencia del evento de interés entre expuestos y no expuestos al factor de riesgo en investigación. Estos estudios a menudo estiman el “riesgo relativo”, el cual puede representar razones de riesgo (razones de incidencias acumuladas) o razones de tasas (razones de tasas incidencia).

En los estudios de casos y controles, sólo se incluye una fracción de la población base (los controles). Los resultados se expresan mediante *odds ratio* de las exposiciones entre casos y controles. Esta *odds ratio* aporta una estimación de la razón de riesgos o la razón de tasas en función del muestreo de casos y controles (véase también el Recuadro 7)^{175,176}. La razón de prevalencias o la *odds ratio* de prevalencia de los estudios transversales pueden ser útiles en algunas situaciones¹⁷⁷.

A menudo puede ser conveniente expresar los resultados tanto en términos relativos como en términos absolutos. Por ejemplo, en un estudio¹⁷⁸ de médicos británicos, la tasa de incidencia de muertes por cáncer de pulmón a lo largo de un seguimiento de 50 años fue de 249 por 100.000 y año entre los fumadores, en comparación con 17 por 100.000 y año entre los no fumadores: una razón de tasas de 14,6 (249/17). Para la enfermedad cardíaca coronaria (ECC), las tasas correspondientes fueron de 1001 y 619 por 100.000 y año, una razón de tasas de 1,61 (1001/619). El efecto del tabaquismo sobre la mortalidad parece mucho más fuerte para el cáncer de pulmón que para la ECC. El cuadro cambia cuando consideramos los efectos absolutos de fumar: la diferencia en las tasas de incidencia fue de 232 por 100.000 y año (249–17) para el cáncer pulmonar y de 382 para la ECC (1001–619). Por lo tanto, entre los médicos que fumaban era más probable que su tabaquismo causara una muerte por ECC que por cáncer de pulmón.

¿Cuánta carga de la enfermedad se podría prevenir en una población eliminando una exposición? Se han publicado estimaciones globales para el caso del tabaquismo: según un

estudio¹⁷⁹, el 91% de todos los cánceres de pulmón, el 40% de las ECC y el 33% de todas las muertes de hombres en el año 2000 se atribuyeron a este factor¹⁷⁹. Generalmente se define la fracción atribuible poblacional como la proporción de casos causados por una exposición particular, pero existen varios conceptos (ninguna terminología unificada), y a veces se utilizan aproximaciones incorrectas para ajustar por otros factores^{172,180}. ¿Cuáles son las implicaciones para los artículos? Las medidas relativas hacen énfasis en la fuerza de una asociación, y son las más útiles de la investigación etiológica. Si se documenta una relación causal con una exposición y las asociaciones se interpretan como efectos, las estimaciones del riesgo relativo se pueden traducir en medidas convenientes del riesgo absoluto con el fin de calcular el posible efecto de las políticas de salud pública (véase el punto 16c)¹⁸¹. Sin embargo, los autores deben ser conscientes de las fuertes suposiciones que se hacen en este contexto¹⁷¹. Se requiere cuidado a la hora de decidir qué concepto y qué método son apropiados para una situación particular.

En muchas circunstancias, el riesgo absoluto asociado a una exposición es de mayor interés que el riesgo relativo. Por ejemplo, si se centra la atención en los efectos nocivos de un fármaco, se deseará saber el número de casos adicionales por unidad de tiempo de uso (por ejemplo, días, semanas o años). El ejemplo presenta el número adicional de casos de cáncer de mama por cada 1000 mujeres que utilizaron terapia hormonal de reemplazo durante 10 años¹⁶³. Medidas como el riesgo o la fracción atribuible poblacional pueden ser útiles para estimar qué proporción de la enfermedad se podría prevenir si se elimina la exposición. Estas medidas se deben presentar preferiblemente junto con una medida de incertidumbre estadística (por ejemplo, intervalos de confianza como en el ejemplo). Los autores deben ser conscientes de las fuertes suposiciones que se hacen en este contexto, incluyendo una relación causal entre un factor de riesgo y una enfermedad (véase también el Recuadro 7)¹⁷¹. Debido a la ambigüedad semántica y a la complejidad que implica, los autores deben indicar detalladamente qué métodos utilizaron para calcular los riesgos atribuibles, idealmente presentando las fórmulas usadas¹⁷².

Un estudio reciente de los resúmenes de 222 artículos publicados en las revistas médicas más importantes encontró que en el 62% de los resúmenes de ensayos aleatorizados que incluyeron una medida de razón se presentaban los riesgos absolutos, en comparación con sólo el 21% de los resúmenes de los estudios de cohortes¹⁷³. Una búsqueda de texto libre en Medline de 1966 a 1997 mostró que 619 artículos mencionaron riesgos atribuibles en el título o en el resumen, en comparación con 18.995 que usaban solamente medidas de riesgo relativo u *odds ratios*, es decir, una proporción de 1 a 31¹⁷⁴.

17. Otros análisis: describa otros análisis efectuados (de subgrupos, interacciones o sensibilidad)

Ejemplo 1

Tabla modificada de Vanderbroucke et al¹⁸² por Botto¹⁸³.

Análisis del uso de anticonceptivos orales				
Factor V de Leiden	Anticonceptivos orales	N.º de pacientes	N.º de controles	Odds ratio
Sí	Sí	25	2	34,7
Sí	No	10	4	6,9
No	Sí	84	63	3,7
No	No	36	63	1
				(referencia)

Presencia del alelo factor V de Leiden y riesgo de tromboembolia venosa.

Ejemplo 2

Tabla adaptada de Wei et al¹⁸⁴.

Sensibilidad de la razón de tasas para un evento cardiovascular con respecto a un confusor no medido

Prevalencia del confusor binario no medido en el grupo expuesto (%)	Prevalencia del confusor binario no medido en el grupo de comparación (%)	Razón de tasas del confusor binario no medido	Razón de tasas para el nivel alto de exposición ^a (IC95%)
90	10	1,5	1,20 (1,01–0,59)
90	50	1,5	1,43 (1,22–1,67)
50	10	1,5	1,39 (1,18–1,63)
90	10	2	0,96 (0,81–1,13)
90	50	2	1,27 (1,11–1,45)
50	10	2	1,21 (1,03–1,42)
90	50	3	1,18 (1,01–1,38)
50	10	3	0,99 (0,85–1,16)
90	50	5	1,08 (0,85–1,26)

IC95%: intervalo de confianza al 95%.

^a Ajustada por edad, sexo, uso de fármaco cardiovascular y confusor binario no medido.

Explicación

Además del análisis principal, en los estudios observacionales a menudo se hacen otros análisis. Pueden referirse a subgrupos específicos, a la posible interacción de factores de riesgo, al cálculo de riesgos atribuibles o al uso de definiciones alternativas de las variables en estudio durante el análisis de sensibilidad.

Hay un debate sobre los peligros asociados al análisis de subgrupos, y en general con la multiplicidad de análisis^{4,104}. En nuestra opinión, también hay una tendencia muy acentuada a buscar evidencias de asociaciones específicas en subgrupos o de modificación del efecto, cuando los resultados globales sugieren poco o ningún efecto. Por otra parte, puede ser valioso explorar si una asociación global parece consistente en varios subgrupos, preferiblemente los preespecificados, en especial cuando un estudio es lo suficiente grande como para tener bastantes datos en cada subgrupo. Un segundo tema de discusión son los subgrupos interesantes que aparecen durante el análisis de datos. Puede que sean hallazgos importantes, pero también puede que aparezcan por casualidad. Algunos argumentan que no es posible ni necesario informar al lector sobre todos los análisis de subgrupos llevados a cabo, ya que análisis futuros de otros datos indicarán en qué medida estos primeros y estimulantes resultados pasan la prueba del tiempo⁹. Nosotros recomendamos a los autores señalar qué análisis se planificaron y cuáles no (véanse también los puntos 4, 12b y 20). Esto permitirá que los lectores juzguen las implicaciones de la multiplicidad, considerando la posición del estudio respecto al continuo que va del descubrimiento a la verificación o a la refutación.

Una tercera cuestión de debate es cómo se deben evaluar los efectos conjuntos y las interacciones de los factores de riesgo: ¿en escalas aditivas o multiplicativas, o la escala se debe determinar por el modelo estadístico que mejor ajuste? (Véanse también el punto 12b y el Recuadro 8.) Una opción apropiada es indicar por separado el efecto de cada exposición, así como el efecto conjunto, si es posible en una tabla, como en el primer ejemplo de arriba¹⁸³ o como en el estudio de Martinelli et al¹⁸⁵. Esta tabla ofrece al lector suficiente información para evaluar tanto la interacción aditiva como la multiplicativa (en el Recuadro 8 se muestra cómo se hacen estos cálculos). Los intervalos de confianza para los efectos aislados y en conjunto pueden ayudar al lector a juzgar la potencia de los datos. Además, los intervalos de confianza de las medidas de interacción, como el exceso del riesgo relativo de la interacción (RERI), se relacionan con las pruebas de interacción o de homogeneidad. Un

Recuadro 8—Interacción (modificación del efecto): el análisis de los efectos conjuntos.

Existe interacción cuando la asociación de una exposición con el riesgo de la enfermedad difiere en presencia de otra exposición. Un problema de la evaluación y la comunicación de interacciones es que el efecto de una exposición se puede medir de dos maneras: como un riesgo relativo (o razón de tasas) o como una diferencia de riesgo (o diferencia de tasas). El uso del riesgo relativo conduce a un modelo multiplicativo, mientras que el uso de la diferencia de riesgo corresponde a un modelo aditivo^{187,188}. A veces se distingue entre la “interacción estadística”, que puede derivarse de un modelo multiplicativo o aditivo, y la “interacción biológica”, que se mide a partir de un modelo aditivo¹⁸⁹. Sin embargo, los modelos aditivos y los multiplicativos no apuntan hacia un mecanismo biológico particular.

Independientemente del modelo elegido, el objetivo principal es entender cómo el efecto conjunto de dos exposiciones difiere de sus efectos separados (en ausencia de la otra exposición). La Red de Epidemiología del Genoma Humano (HuGENet) propuso un esquema para la presentación transparente de efectos separados y conjuntos, que permite evaluar diversos tipos de interacción¹⁸³. Se emplearon datos del estudio sobre anticonceptivos orales y mutación del factor V de Leiden¹⁸² para explicar la propuesta, y este ejemplo también se ha empleado en el punto 17. Los anticonceptivos orales y la mutación del factor V de Leiden incrementan el riesgo de trombosis venosa; sus efectos separados y conjuntos se pueden calcular de la tabla 2 × 4 (véase el ejemplo 1 del punto 17), donde la *odds ratio* de 1 denota la línea basal de las mujeres sin el factor V de Leiden que no utilizan anticonceptivos orales.

Una dificultad es que los diseños de algunos estudios, como los de casos y controles, y varios modelos estadísticos, como los logísticos o de regresión de Cox, estiman riesgos relativos (o razones de tasas) e intrínsecamente conducen a un modelo multiplicativo. En estos casos, los riesgos relativos se pueden traducir a una escala aditiva. En el ejemplo 1 del punto 17, las *odds ratios* separadas son 3,7 y 6,9, mientras que la conjunta es 34,7. Cuando estos datos se analizan con un modelo multiplicativo, se espera una *odds ratio* conjunta de 25,7 (3,7 × 6,9). El efecto conjunto observado, de 34,7, es 1,4 mayor que el esperado en una escala multiplicativa (34,7/25,7). Esta cantidad (1,4) es la *odds ratio* de la interacción multiplicativa. Sería igual al antilogaritmo del coeficiente estimado de la interacción de un modelo de regresión logística. Con el modelo aditivo se espera que la *odds ratio* conjunta sea 9,6 (3,7+6,9–1). El efecto conjunto observado se aleja fuertemente de la aditividad: la diferencia es 25,1 (34,7–9,6). Cuando las *odds ratios* se interpretan como riesgos relativos (o razones de tasas), la última cantidad (25,1) es el exceso del riesgo relativo de la interacción (RERI)¹⁹⁰. Esto se puede entender más fácilmente al imaginarse que el valor de referencia (equivalente a una *odds ratio* = 1) representa una incidencia basal de trombosis venosa de, por ejemplo, 1/10.000 mujeres-año, que aumenta en presencia de exposiciones separadas y conjuntas.

problema recurrente es que los autores utilizan comparaciones de valores p entre subgrupos, lo que lleva a suponer erróneamente una modificación del efecto. Por ejemplo, una asociación estadísticamente significativa en una categoría (hombres), pero no en la otra (mujeres), en sí misma no proporciona evidencia de la modificación del efecto. Del mismo modo, los intervalos de confianza para cada estimador concreto se utilizan a veces de forma inadecuada para inferir que no hay interacción cuando los intervalos se solapan. Se consigue una inferencia más válida evaluando directamente si la magnitud de una asociación difiere entre los subgrupos.

Los análisis de sensibilidad son útiles para investigar la influencia de las decisiones que se toman durante el análisis estadístico, y para investigar la robustez de los hallazgos con respecto a los datos ausentes o a los posibles sesgos (véase también el punto 12b). Es necesario tener un criterio explícito acerca de la cantidad de información sobre tales análisis. Si se realizaron muchos análisis de sensibilidad, puede ser poco práctico presentar los resultados detallados de todos ellos. A veces puede ser suficiente señalar que se realizaron los análisis de sensibilidad y que fueron consistentes con los principales resultados presentados. La presentación detallada es más apropiada si el problema investigado es importante o si las estimaciones del efecto varían considerablemente^{59,186}.

Pocock et al.⁴ encontraron que 43 de 73 artículos que presentaban estudios observacionales contenían análisis de subgrupos. La mayoría señalaba diferencias entre grupos, pero solamente ocho artículos presentaban una evaluación formal de la interacción (véase el punto 12(b)).

Discusión

La sección de la discusión trata los temas centrales de la validez y el significado del estudio¹⁹¹. Algunos trabajos han encontrado que las secciones de la discusión a menudo están dominadas por evaluaciones incompletas o sesgadas de los resultados del estudio y de sus implicaciones, así como con retórica que apoya los resultados de los autores^{192,193}. Estructurar la discusión puede ayudar a los autores a evitar la especulación y la sobreinterpretación injustificadas de los resultados, y al mismo tiempo orientar a los lectores a lo largo del texto^{194,195}. Por ejemplo, *Annals of Internal Medicine*¹⁹⁶ recomienda que los autores estructuren la sección de la discusión presentando lo siguiente: 1) una breve sinopsis de los resultados clave; 2) una consideración de los posibles mecanismos y explicaciones; 3) una comparación con los resultados relevantes de otros estudios publicados; 4) las limitaciones del estudio; y 5) un breve apartado que resuma las implicaciones del trabajo para la práctica y la investigación. Otros han hecho sugerencias similares^{191,194}. Los apartados sobre recomendaciones para la investigación y sobre las limitaciones del estudio deben relacionarse de forma estrecha uno con otro. Los investigadores deben sugerir modos de mejorar la investigación subsecuente más que simplemente señalar que “se requiere más investigación”^{197,198}. Recomendamos que los autores estructuren sus apartados de la discusión, incluso empleando subtítulos convenientes.

18. Resultados clave: resuma los resultados principales de los objetivos del estudio

Ejemplo

“Nuestra hipótesis era que pertenecer a una etnia minoritaria estaría asociado con un mayor riesgo de enfermedad cardiovascular (ECV), pero que la asociación se explicaría sustancialmente por el nivel socioeconómico (NSE). Esta hipótesis no se confirmó. Después de ajustar por la edad y por NSE, siguieron existiendo diferencias altamente significativas en el índice de masa corporal, la presión sanguínea, la diabetes y la inactividad física entre las mujeres blancas y las mujeres de raza negra, así como con las mujeres mexicanas-americanas. Además, encontramos grandes diferencias en los factores de riesgo para ECV por NSE, un hallazgo que ilustra la situación de alto riesgo tanto de las mujeres en la etnia minoritaria como de las mujeres blancas con NSE bajo”¹⁹⁹.

Explicación

Es una buena práctica comenzar la discusión con un resumen breve de los principales hallazgos del estudio. Un resumen corto

recuerda a los lectores los hallazgos principales y les ayuda a evaluar si las interpretaciones subsecuentes y las implicaciones mostradas por los autores están apoyadas por los hallazgos.

19. Limitaciones: discuta las limitaciones del estudio, teniendo en cuenta posibles fuentes de sesgo o de imprecisión. razone tanto sobre la dirección como sobre la magnitud de cualquier posible sesgo

Ejemplo

“Dado que la frecuencia del consejo aumenta con el mayor grado de obesidad, nuestros estimadores pueden sobrestimar la verdadera prevalencia de la búsqueda de ayuda. Las encuestas telefónicas también pueden sobrestimar la verdadera prevalencia del consejo. Aunque las personas sin teléfono tienen similares grados de sobrepeso que las personas con teléfono, aquéllas sin este servicio tienden a ser menos educadas, un factor asociado con menos consejo en nuestro estudio. También llama la atención el posible sesgo causado por los que rechazaron participar, así como por los que rechazaron responder a las preguntas acerca del peso corporal. Además, debido a que los datos se recogieron transversalmente, no podemos inferir que el consejo precediera al intento del paciente de perder peso”²⁰⁰.

Explicación

La identificación y la discusión de las limitaciones de un estudio son una parte esencial del artículo científico. Es importante no sólo identificar el origen de los sesgos y de la confusión que pueden afectar los resultados, sino también comentar la importancia relativa de los diferentes sesgos, incluyendo la dirección y la magnitud de cualquier posible sesgo (véanse el [Recuadro 3](#) y el punto 9).

Los autores deben discutir también cualquier imprecisión de los resultados. Esta imprecisión puede surgir respecto a varios aspectos del estudio, incluyendo su tamaño (punto 10) y la medida de la exposición, los confusores y los resultados (punto 8). La incapacidad para medir con precisión los verdaderos valores de una exposición tiende a derivar en un sesgo hacia la unidad: cuanto menor sea la precisión en la medida del factor de riesgo, mayor será el sesgo. Este efecto se ha descrito como “atenuación”^{201,202} y más recientemente como “sesgo de dilución por la regresión”²⁰³. Sin embargo, cuando los factores de riesgo correlacionados se miden con diferentes niveles de imprecisión, el riesgo relativo ajustado asociado a ellos puede estar sesgado en la dirección o en contra de la unidad^{204–206}.

Al comentar las limitaciones, los autores pueden comparar su estudio con otros de la literatura en términos de validez, generalización y precisión. Con este enfoque, cada estudio se puede ver como una contribución a la literatura, no como una base autónoma para la inferencia y la acción²⁰⁷. Sorprendentemente, la discusión de las limitaciones importantes de un estudio a veces se omite en los artículos publicados. Una encuesta entre los autores que habían publicado artículos de investigación original en la revista *The Lancet* encontró que los investigadores señalaban las debilidades importantes de los estudios en la encuesta, pero no en los artículos que habían publicado¹⁹².

20. Interpretación: proporcione una interpretación global prudente de los resultados considerando objetivos, limitaciones, multiplicidad de análisis, resultados de estudios similares y otras pruebas empíricas relevantes

Ejemplo

“Cualquier explicación de una asociación entre muerte por infarto de miocardio y uso de anticonceptivos orales de segunda

generación debe ser una conjetura. No hay evidencia publicada que sugiera un mecanismo biológico directo, y no hay otros estudios epidemiológicos con resultados relevantes. (...) El incremento en el riesgo absoluto es muy pequeño y probablemente se aplique predominantemente a fumadores. Debido a la falta de evidencia que lo confirme, y considerando que el análisis se basó en cifras relativamente pequeñas, se necesitan más pruebas al respecto. No recomendaríamos ningún cambio en la práctica de prescripción basándose en estos resultados”¹²⁰.

Explicación

El núcleo de la sección de la discusión es la interpretación de los resultados del estudio. La sobreinterpretación es común y de humanos; aun cuando nos esforzamos por ofrecer una evaluación objetiva, con frecuencia los revisores nos señalan que fuimos muy lejos en algunos aspectos. Cuando se interpreten los resultados, los autores deben considerar la naturaleza del estudio basándose en la verificación continua de los descubrimientos y en las posibles fuentes de sesgo, incluyendo las pérdidas durante el seguimiento y la no participación (ver puntos 9, 12 y 19). Se deben considerar adecuadamente el tema de la confusión (punto 16a) y los resultados relevantes del análisis de sensibilidad, así como el problema de la multiplicidad y el análisis de los diferentes subgrupos (punto 17). Los autores siempre deben tener en cuenta la confusión residual debida a variables no medidas o a una determinación imprecisa de los confusores. Por ejemplo, el nivel socioeconómico se asocia con varios resultados en salud y a menudo difiere entre los grupos que se comparan. Las variables que se utilizan para medir el nivel socioeconómico (ingresos, educación u ocupación) representan otras exposiciones no medidas y no definidas, y el verdadero confusor será medido, por definición, con error²⁰⁸. Los autores deben discutir el rango real de incertidumbre en los estimadores, el cual es mayor que la incertidumbre estadística reflejada en los intervalos de confianza. Esto último no toma en consideración otras incertidumbres que surgen del diseño, la implementación y los métodos de medida del estudio²⁰⁹.

Para guiar el juicio y las conclusiones respecto a la causalidad, algunos pueden encontrar útiles los criterios propuestos por Bradford Hill¹⁶⁴ en 1965. ¿Cuán fuerte es la asociación con la exposición? ¿Precedió la exposición al inicio de la enfermedad? ¿La asociación es consistente con observaciones en diferentes estudios y contextos? ¿Hay evidencia que apoye los hallazgos provenientes de estudios experimentales, incluyendo estudios de laboratorio y en animales? ¿Cuán específica es la exposición para el efecto que se le imputa y existe una relación dosis-respuesta? ¿La asociación es biológicamente plausible? No obstante, estos criterios no se deben aplicar mecánicamente. Por ejemplo, algunos han argumentado que un riesgo relativo por debajo de 2 o 3 se debe ignorar^{210,211}. Esto es contrario a lo expuesto por Cornfield et al¹²⁷ acerca de la fuerza de los riesgos relativos grandes (ver punto 12b). Aunque un efecto causal es más probable con un riesgo relativo de 9, no se puede deducir que uno por debajo de 3 sea necesariamente falso. Por ejemplo, el pequeño incremento en el riesgo de presentar leucemia en la niñez después de haber recibido radiaciones intraútero es verosímil porque tiene que ver con un efecto adverso de un procedimiento médico, para el cual ninguna explicación alternativa es obvia²¹². Además, los efectos cancerígenos de la radiación están bien establecidos. El doble del riesgo de cáncer de ovario asociado a comer 2 o 4 huevos por semana no es creíble inmediatamente, puesto que los hábitos alimentarios se asocian con un gran número de factores del estilo de vida, así como con el nivel socioeconómico²¹³. En contraste, la credibilidad de los ampliamente debatidos hallazgos epidemiológicos respecto a las diferencias en el riesgo de trombosis entre los

distintos tipos de anticonceptivos orales aumentó enormemente por las diferencias en la coagulación que se encontraron en un estudio clínico cruzado y aleatorizado²¹⁴. Siempre se debe incluir una discusión de la evidencia externa de diferentes tipos de estudios, pero esto puede ser particularmente importante para aquellos que comuniquen pequeños incrementos en el riesgo. Además, los autores deben contextualizar sus resultados con estudios similares y explicar cómo el nuevo estudio afecta el cuerpo de evidencia existente, idealmente refiriéndose a una revisión sistemática.

21. Generabilidad: discuta la posibilidad de generalizar los resultados (validez externa)

Ejemplo

“¿Son aplicables nuestras estimaciones a otros pacientes infectados por el VIH-1? Ésta es una pregunta importante, porque la precisión de los modelos pronóstico tiende a ser menor cuando se aplican datos distintos a los utilizados para el modelo. Nosotros tratamos este problema simplificando la complejidad de los modelos y escogiendo aquellos que mejor se generalizaran para las cohortes omitidas en el procedimiento de estimación. Nuestra base de datos incluyó pacientes de muchos países de Europa y Norteamérica que fueron tratados en diferentes contextos. La diversidad de pacientes fue amplia: se incluyeron hombres y mujeres, desde adolescentes hasta personas de edad avanzada, y las principales categorías de exposición estuvieron bien representadas. La gravedad de la inmunodeficiencia basal varió desde no medible hasta muy grave, y la carga viral desde indetectable hasta extremadamente alta”²¹⁵.

Explicación

La generabilidad, también llamada validez externa o aplicabilidad, se refiere al grado en que los resultados de un estudio se pueden aplicar a otras circunstancias²¹⁶. No hay validez externa *per se*; el término sólo es significativo con respecto a condiciones específicas muy claras²¹⁷. ¿Se pueden aplicar los resultados a individuos, grupos o poblaciones que difieren de aquellos que participaron en el estudio en lo que se refiere a edad, sexo, etnia, gravedad de la enfermedad y condiciones de comorbilidad? ¿La naturaleza y el grado de las exposiciones son comparables, y la definición de los eventos de interés es relevante para otros contextos o poblaciones? ¿Los datos recogidos en estudios longitudinales varios años atrás siguen siendo relevantes hoy día? ¿Los resultados de las investigaciones sobre servicios de salud en un país son aplicables a los sistemas de salud de otros países?

Con frecuencia, la pregunta sobre si los resultados de un estudio tienen validez externa es una cuestión de juicio que depende del contexto del estudio, de las características de los participantes, de las exposiciones estudiadas y de los desenlaces evaluados. Por lo tanto, es crucial que los autores presenten a los lectores información adecuada respecto a las condiciones y el lugar del estudio, los criterios de elegibilidad, las exposiciones y cómo se midieron, la definición del desenlace evaluado, así como el periodo de reclutamiento y de seguimiento. También son relevantes el grado de no participación y la proporción de participantes no expuestos en que se desarrolló el resultado. El conocimiento del riesgo absoluto y de la prevalencia de la exposición, que a menudo varían en las diferentes poblaciones, son de ayuda cuando se aplican resultados a otros ámbitos o poblaciones (véase el Recuadro 7).

Otra información

22. *Financiación: especifique la financiación y el papel de los patrocinadores del estudio, y si procede, del estudio previo en que se basa su artículo*

Explicación

Algunas publicaciones solicitan a los autores que declaren la presencia o ausencia de financiación y otros conflictos de intereses^{100,218}. Varias investigaciones muestran fuertes asociaciones entre el origen de la financiación y las conclusiones de los artículos de investigación^{219–222}. Las conclusiones de algunos estudios clínicos aleatorizados han recomendado el fármaco en estudio como de elección con mucha más frecuencia (*odds ratio* de 5,3) si el estudio estaba financiado por organizaciones lucrativas, aun después de ajustar por el tamaño del efecto²²³. Otros estudios documentan la influencia de las industrias del tabaco y de telecomunicaciones en las investigaciones que éstas financian^{224–227}. También hay ejemplos de influencia indebida cuando el patrocinador es el gobierno u organizaciones no lucrativas.

Los autores o patrocinadores pueden tener conflictos de intereses que influyan en cualquiera de los siguientes aspectos: el diseño del estudio²²⁸, la selección de las exposiciones^{228,229}, los desenlaces²³⁰, los métodos estadísticos²³¹ y la publicación selectiva de los resultados²³⁰ y de los estudios²³². En consecuencia, el papel de los patrocinadores se debe describir con detalle: en qué parte del estudio tienen responsabilidad directa (por ejemplo, el diseño, la recogida de los datos, los análisis, el borrador del manuscrito, la decisión de publicar)¹⁰⁰. Otras fuentes de influencia indebida incluyen a los empleadores (por ejemplo, los administradores de la universidad en el caso de los investigadores académicos y de los supervisores del gobierno, especialmente los asignados por motivos políticos, para los investigadores gubernamentales), comités asesores, litigantes y grupos de interés especiales.

Comentarios finales

La Declaración STROBE tiene como fin ofrecer recomendaciones que ayuden a la comunicación de estudios observacionales en epidemiología. Un buen artículo revela las fortalezas y las debilidades de un estudio, y facilita las interpretaciones sensatas y la aplicación de sus resultados. La Declaración STROBE también puede ayudar en la planificación de los estudios observacionales, y guiar a los revisores y editores en la evaluación de los manuscritos.

Este artículo explicativo se escribió para comentar la importancia de un informe transparente y completo de los estudios observacionales, para explicar el fundamento que respalda los diferentes puntos incluidos en la lista y para dar ejemplos de artículos publicados que consideramos bien realizados. Esperamos que el material presentado ayude a los autores y editores a utilizar la Declaración STROBE.

Hacemos hincapié en que la Declaración STROBE y otras recomendaciones sobre la comunicación de investigaciones^{13,233,234} deben ser vistas como documentos que han de evolucionar, que requieren una evaluación, un refinamiento y, si es necesario, un cambio continuo^{235,236}. Por ejemplo, la Declaración CONSORT²³⁷ para la publicación de estudios clínicos paralelos aleatorizados se desarrolló a mediados de los años 1990. Desde entonces, los miembros del grupo tienen reuniones regulares para examinar la necesidad de revisar las recomendaciones; en 2001 apareció una versión revisada²³³ y ya hay otra en desarrollo. Igualmente, los principios presentados en este artículo y en la lista de la Declaración STROBE están abiertos a

modificaciones en cuanto se acumulen nuevas evidencias y comentarios críticos. El sitio web de STROBE (www.strobe-statement.org) proporciona un foro de discusión y sugerencias para mejorar la lista, este documento explicativo y la información relacionada con la buena comunicación de los estudios epidemiológicos.

Varias revistas ya solicitan en sus instrucciones para los autores que se siga la Declaración STROBE (véase <http://www.strobe-statement.org/> para la lista actual). Invitamos a otras revistas a adoptar la Declaración STROBE y a que nos lo hagan saber a través del sitio web. Las revistas que publicaron las recomendaciones STROBE dan acceso libre. Por lo tanto, la Declaración STROBE es ampliamente accesible para la comunidad biomédica.

Agradecimientos

Estamos agradecidos a Gerd Antes, Kay Dickersin, Shah Ebrahim y Richard Lilford por su apoyo a la Iniciativa STROBE. También queremos dar las gracias a las instituciones que han ofrecido sus instalaciones para las reuniones de trabajo del grupo de coordinación: Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Suiza; Department of Social Medicine, University of Bristol, Bristol, Reino Unido; London School of Hygiene and Tropical Medicine, Londres, Reino Unido; Nordic Cochrane Centre, Copenhage, Dinamarca, y Centre for Statistics in Medicine, Oxford, Reino Unido. También queremos expresar nuestro agradecimiento a seis revisores que aportaron comentarios útiles a una versión previa de este artículo y al traductor y revisores de la traducción al español.

Financiación

El seminario en el que se estableció la Iniciativa STROBE fue financiado por la European Science Foundation (ESF). Se recibió también financiación del Medical Research Council Health Services Research Collaboration y del National Health Services Research and Development Methodology Programme.

Contribuciones de autoría

Todos los autores contribuyeron en la redacción del artículo. JPV, EvE, DGA, PCG, SJP y ME escribieron el primer borrador de diferentes secciones del artículo. EvE asume la mayor parte de las cuestiones prácticas relativas a la coordinación de STROBE. ME inició el grupo STROBE y, junto con EvE, organizó el primer seminario.

Miembros de la Iniciativa STROBE

Las siguientes personas han contribuido a los contenidos y la elaboración de la Declaración STROBE: Douglas G. Altman, Maria Blettner, Paolo Boffetta, Hermann Brenner, Geneviève Chêne, Cyrus Cooper, George Davey-Smith, Erik von Elm, Matthias Egger, France Gagnon, Peter C. Gøtzsche, Philip Greenland, Sander Greenland, Claire Infante-Rivard, John Ioannidis, Astrid James, Giselle Jones, Bruno Ledergerber, Julian Little, Margaret May, David Moher, Hooman Momen, Alfredo Morabia, Hal Morgenstern, Cynthia D. Mulrow, Fred Paccaud, Stuart J. Pocock, Charles Poole, Martin Rössli, Drummond Rennie, Dietrich Rothenbacher, Kenneth Rothman, Caroline Sabin, Willi Sauerbrei, Lale Say, James J. Schlesselman, Jonathan Sterne, Holly Syddall, Jan P. Vandenbroucke, Ian White, Susan Wieland, Hywel Williams y Guang Yong Zou.

Declaración de conflicto de intereses

Los autores declaran que no existen conflictos de interés.

Copyright: © 2007 Los Autores. Hay más detalles sobre el uso de este material en la página web STROBE (<http://www.strobe-statement.org>). Con el objetivo de fomentar la divulgación de la Declaración STROBE, este artículo también se ha publicado y está accesible gratuitamente en las revistas PLoS Medicine, Epidemiology y Annals of Internal Medicine.

La versión en español ha sido traducida del artículo publicado en PLoS Medicine (PLoS Med 4(10): e297) y ha sido autorizada y aprobada por el grupo coordinador de la Iniciativa STROBE. El responsable de la traducción es Mario E. Rojas-Russell y han colaborado M^a de los Ángeles Aedo-Santos, Hernán Chivardi-Olmos, Elsy A. García-Villegas, Eloísa Colín-Ramírez, Reyna L. Pacheco-Domínguez, Manuel C. Ortega-Álvarez, todos ellos de la Facultad de Medicina de la Universidad Nacional Autónoma de México. Los traductores agradecen a Miguel Klunder-Klunder y Clara Bellamy-Ortiz sus comentarios al texto. La traducción ha sido supervisada por Malaquías López-Cervantes, Carmen Company y Esteve Fernández.

Bibliografía

- Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ*. 2004;328:39–41.
- Funai EF, Rosenbush EJ, Lee MJ, Del Priore G. Distribution of study designs in four major US journals of obstetrics and gynecology. *Gynecol Obstet Invest*. 2001;51:8–11.
- Scales Jr CD, Norris RD, Peterson BL, et al. Clinical research and statistical methods in the urology literature. *J Urol*. 2005;174:1374–9.
- Pocock SJ, Collier TJ, Dandreo KJ, et al. Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ*. 2004;329:883.
- Tooth L, Ware R, Bain C, et al. Quality of reporting of observational longitudinal research. *Am J Epidemiol*. 2005;161:280–8.
- Von Elm E, Altman DG, Egger M, et al. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *PLoS Med* e296. 2007; doi:10.1371/journal.pmed.0040296.
- Mihailovic A, Bell CM, Urbach DR. Users' guide to the surgical literature. Case-control studies in surgical journals. *Can J Surg*. 2005;48:148–51.
- Rushton L. Reporting of occupational and environmental research: use and misuse of statistical and epidemiological methods. *Occup Environ Med*. 2000;57:1–9.
- Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology*. 1990;1:43–6.
- Moonesinghe R, Khoury MJ, Janssens AC. Most published research findings are false but a little replication goes a long way. *PLoS Med* 4. 2007:e28.
- Jenicek M. Clinical case reporting. Evidence-based medicine. Oxford: Butterworth-Heinemann; 1999.
- Vandenbroucke JP. In defense of case reports and case series. *Ann Intern Med*. 2001;134:330–4.
- Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD Initiative. *Ann Intern Med*. 2003;138:40–4.
- McShane LM, Altman DG, Sauerbrei W, et al. Reporting recommendations for tumor marker prognostic studies (REMARK). *Br J Cancer*. 2005;93:387–91.
- Ioannidis JP, Gwinn M, Little J, et al. A road map for efficient and reliable human genome epidemiology. *Nat Genet*. 2006;38:3–5.
- Rodríguez L, Kirkwood BR. Case-control designs in the study of common diseases: updates on the demise of the rare disease assumption and the choice of sampling scheme for controls. *Int J Epidemiol*. 1990;19:205–13.
- Rothman KJ, Greenland S. Case-control studies. En: Rothman KJ, Greenland S, editores. *Modern epidemiology*. 2nd ed. Filadelfia: Lippincott Raven; 1998. p. 93–114.
- Forand SP. Leukemia incidence among workers in the shoe and boot manufacturing industry: a case-control study. *Environ Health*. 2004;3:7.
- Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med*. 2000;342:1878–86.
- Gotzsche PC, Harden A. Searching for non-randomised studies. Draft chapter 3. Cochrane Non-Randomized Studies Methods Group, 26 julio 2002. Disponible en: <http://www.cochrane.dk/nrsmg> (acceso el 10 de septiembre de 2007).
- Lohse N, Hansen AB, Pedersen G, et al. Survival of persons with and without HIV infection in Denmark, 1995–2005. *Ann Intern Med*. 2007;146:87–95.
- American Journal of Epidemiology. Information for authors. Disponible en: http://www.oxfordjournals.org/aje/for_authors/index.html (acceso el 10 de septiembre de 2007).
- Haynes RB, Mulrow CD, Huth EJ, et al. More informative abstracts revisited. *Ann Intern Med*. 1990;113:69–76.
- Taddio A, Pain T, Fassos FF, et al. Quality of nonstructured and structured abstracts of original research articles in the British Medical Journal, the Canadian Medical Association Journal and the Journal of the American Medical Association. *CMAJ*. 1994;150:1611–5.
- Hartley J, Sydes M. Which layout do you prefer? An analysis of readers' preferences for different typographic layouts of structured abstracts. *J Inform Sci*. 1996;22:27–37.
- Viner RM, Cole TJ. Adult socioeconomic, educational, social, and psychological outcomes of childhood obesity: a national birth cohort study. *BMJ*. 2005;330:1354.
- McCauley J, Kern DE, Kolodner K, et al. The "battering syndrome": prevalence and clinical characteristics of domestic violence in primary care internal medicine practices. *Ann Intern Med*. 1995;123:737–46.
- McEvoy SP, Stevenson MR, McCart AT, et al. Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. *BMJ*. 2005;331:428.
- Vandenbroucke JP. Prospective or retrospective: what's in a name? *BMJ*. 1991;302:249–50.
- Last JM. A dictionary of epidemiology. New York: Oxford University Press; 2000.
- Miettinen OS. Theoretical epidemiology: principles of occurrence research in medicine. New York: Wiley; 1985. p. 64–6.
- Rothman KJ, Greenland S. Types of epidemiologic studies. En: Rothman KJ, Greenland S, editores. *Modern epidemiology*. 2nd ed. Boston: Lippincott Raven; 1998. p. 74–5.
- MacMahon B, Trichopoulos D. *Epidemiology, principles and methods*. 2nd ed. Boston: Little, Brown; 1996. p. 81.
- Lilienfeld AM. *Foundations of epidemiology*. New York: Oxford University Press; 1976.
- Ridker PM, Hennekens CH, Lindpaintner K, et al. Mutation in the gene coding for coagulation factor V and the risk of myocardial infarction, stroke, and venous thrombosis in apparently healthy men. *N Engl J Med*. 1995;332:912–7.
- Goodman KJ, O'Rourke K, Day RS, et al. Dynamics of Helicobacter pylori infection in a US-Mexico cohort during the first two years of life. *Int J Epidemiol*. 2005;34:1348–55.
- Altman DG, De Stavola BL, Love SB, et al. Review of survival analyses published in cancer journals. *Br J Cancer*. 1995;72:511–8.
- Cerhan JR, Wallace RB, Folsom AR, et al. Transfusion history and cancer risk in older women. *Ann Intern Med*. 1993;119:8–15.
- Beane Freeman LE, Dennis LK, Lynch CF, et al. Toenail arsenic content and cutaneous melanoma in Iowa. *Am J Epidemiol*. 2004;160:679–87.
- Canto JG, Allison JJ, Kiefe CI, et al. Relation of race and sex to the use of reperfusion therapy in Medicare beneficiaries with acute myocardial infarction. *N Engl J Med*. 2000;342:1094–100.
- Metzker-Cotter E, Kletter Y, Avidor B, et al. Long-term serological analysis and clinical follow-up of patients with cat scratch disease. *Clin Infect Dis*. 2003;37:1149–54.
- Johnson ES. Bias on withdrawing lost subjects from the analysis at the time of loss, in cohort mortality studies, and in follow-up methods. *J Occup Med*. 1990;32:250–4.
- Berkson J. Limitations of the application of fourfold table analysis to hospital data. *Biom Bull*. 1946;2:53.
- Feinstein AR, Walter SD, Horwitz RJ. An analysis of Berkson's bias in case-control studies. *J Chronic Dis*. 1986;39:495–504.
- Jick H, Vessey MP. Case-control studies in the evaluation of drug induced illness. *Am J Epidemiol*. 1978;107:1–7.
- Hackam DG, Mamdani M, Li P, Redelmeier DA. Statins and sepsis in patients with cardiovascular disease: a population-based cohort analysis. *Lancet*. 2006;367:413–8.
- Smeeth L, Cook C, Fombonne E, et al. MMR vaccination and pervasive developmental disorders: a case-control study. *Lancet*. 2004;364:963–9.
- Costanza MC. Matching. *Prev Med*. 1995;24:425–33.
- Sturmer T, Brenner H. Flexible matching strategies to increase power and efficiency to detect and estimate gene-environment interactions in case-control studies. *Am J Epidemiol*. 2002;155:593–602.
- Rothman KJ, Greenland S. Matching. En: Rothman KJ, Greenland S, editores. 2nd ed. *Modern epidemiology*. Boston: Lippincott Raven; 1998. p. 147–61.
- Szklo MF, Nieto J. *Epidemiology, beyond the basics*. Sudbury (MA): Jones and Bartlett; 2000. p. 40–51.
- Cole P, MacMahon B. Attributable risk percent in case-control studies. *Br J Prev Soc Med*. 1971;25:242–4.
- Gissler M, Hemminki E. The danger of overmatching in studies of the perinatal mortality and birthweight of infants born after assisted conception. *Eur J Obstet Gynecol Reprod Biol*. 1996;69:73–5.
- Gefeller O, Pfahlberg A, Brenner H, Windeler J. An empirical investigation on matching in published case-control studies. *Eur J Epidemiol*. 1998;14:321–5.
- Artama M, Ritvanen A, Gissler M, et al. Congenital structural anomalies in offspring of women with epilepsy—a population based cohort study in Finland. *Int J Epidemiol*. 2006;35:280–7.
- Ebrahim S. Cohorts, infants and children. *Int J Epidemiol*. 2004;33:1165–6.
- Walker M, Whincup PH, Shaper AG. The British Regional Heart Study 1975–2004. *Int J Epidemiol*. 2004;33:1185–92.
- Wieland S, Dickersin K. Selective exposure reporting and Medline indexing limited the search sensitivity for observational studies of the adverse effects of oral contraceptives. *J Clin Epidemiol*. 2005;58:560–7.
- Anderson HR, Atkinson RW, Peacock JL, et al. Ambient particulate matter and health effects: publication bias in studies of short-term associations. *Epidemiology*. 2005;16:155–63.

60. Winkelmayr WC, Stampfer MJ, Willett WC, et al. Habitual caffeine intake and the risk of hypertension in women. *JAMA*. 2005;294:2330–5.
61. Lukanova A, Soderberg S, Kaaks R, et al. Serum adiponectin is not associated with risk of colorectal cancer. *Cancer Epidemiol Biomarkers Prev*. 2006;15:401–2.
62. Becher H. The concept of residual confounding in regression models and some applications. *Stat Med*. 1992;11:1747–58.
63. Brenner H, Blettner M. Controlling for continuous confounders in epidemiologic research. *Epidemiology*. 1997;8:429–34.
64. Phillips MR, Yang G, Zhang Y, et al. Risk factors for suicide in China: a national case-control psychological autopsy study. *Lancet*. 2002;360:1728–36.
65. Pasquale LR, Kang JH, Manson JE, et al. Prospective study of type 2 diabetes mellitus and risk of primary open angle glaucoma in women. *Ophthalmology*. 2006;113:1081–6.
66. Craig SL, Feinstein AR. Antecedent therapy versus detection bias as causes of neoplastic multimorbidity. *Am J Clin Oncol*. 1999;22:51–6.
67. Rogler LH, Mroczek DK, Fellows M, et al. The neglect of response bias in mental health research. *J Nerv Ment Dis*. 2001;189:182–7.
68. Murphy EA. *The logic of medicine*. Baltimore: Johns Hopkins University Press; 1976.
69. Sackett DL. Bias in analytic research. *J Chronic Dis*. 1979;32:51–63.
70. Johannes CB, Crawford SL, McKinlay JB. Interviewer effects in a cohort study. Results from the Massachusetts Women's Health Study. *Am J Epidemiol*. 1997;146:429–38.
71. Bloemenkamp KW, Rosendaal FR, Buller HR, et al. Risk of venous thrombosis with use of current low-dose oral contraceptives is not explained by diagnostic suspicion and referral bias. *Arch Intern Med*. 1999;159:65–70.
72. Feinstein AR. *Clinical epidemiology: the architecture of clinical research*. Philadelphia: WB Saunders; 1985.
73. Yadon ZE, Rodrigues LC, Davies CR, et al. Indoor and peridomestic transmission of American cutaneous leishmaniasis in northwestern Argentina: a retrospective case-control study. *Am J Trop Med Hyg*. 2003;68:519–26.
74. Anoop S, Saravanan B, Joseph A, et al. Maternal depression and low maternal intelligence as risk factors for malnutrition in children: a community based case-control study from South India. *Arch Dis Child*. 2004;89:325–9.
75. Carlin JB, Doyle LW. Sample size. *J Paediatr Child Health*. 2002;38:300–4.
76. Rigby AS, Vail A. Statistical methods in epidemiology. II: a common-sense approach to sample size estimation. *Disabil Rehabil*. 1998;20:405–10.
77. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet*. 2005;365:1348–53.
78. Drescher K, Timm J, Jockel KH. The design of case-control studies: the effect of confounding on sample size requirements. *Stat Med*. 1990;9:765–76.
79. Devine OJ, Smith JM. Estimating sample size for epidemiologic studies: the impact of ignoring exposure measurement uncertainty. *Stat Med*. 1998;17:1375–89.
80. Linn S, Levi L, Grunau PD, et al. Effect measure modification and confounding of severe head injury mortality by age and multiple organ injury severity. *Ann Epidemiol*. 2007;17:142–7.
81. Altman DG, Lausen B, Sauerbrei W, et al. Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst*. 1994;86:829–35.
82. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*. 2006;25:127–41.
83. Greenland S. Avoiding power loss associated with categorization and ordinal scores in dose-response and trend analysis. *Epidemiology*. 1995;6:450–4.
84. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol*. 1999;28:964–74.
85. MacCallum RC, Zhang S, Preacher KJ, et al. On the practice of dichotomization of quantitative variables. *Psychol Methods*. 2002;7:19–40.
86. Altman DG. Categorizing continuous variables. En: Armitage P, Colton T, editores. *Encyclopedia of biostatistics*. 2nd ed. Chichester: John Wiley; 2005. p. 708–11.
87. Cohen J. The cost of dichotomization. *Applied Psychological Measurement*. 1983;7:249–53.
88. Zhao LP, Kolonel LN. Efficiency loss from categorizing quantitative exposures into qualitative exposures in case-control studies. *Am J Epidemiol*. 1992;136:464–74.
89. Cochran WG. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*. 1968;24:295–313.
90. Clayton D, Hills M. *Statistical models in epidemiology*. Oxford: Oxford University Press; 1993. p. 249–60.
91. Cox DR. Note on grouping. *J Am Stat Assoc*. 1957;52:543–7.
92. Il'yasova D, Hertz-Picciotto I, Peters U, et al. Choice of exposure scores for categorical regression in meta-analysis: a case study of a common problem. *Cancer Causes Control*. 2005;16:383–8.
93. Berglund A, Alfredsson L, Cassidy JD, et al. The association between exposure to a rear-end collision and future neck or shoulder pain: a cohort study. *J Clin Epidemiol*. 2000;53:1089–94.
94. Slama R, Werwatz A. Controlling for continuous confounding factors: non and semiparametric approaches. *Rev Epidemiol Sante Publique*. 2005;53(Spec No 2): 2565–80.
95. Greenland S. Introduction to regression modelling. En: Rothman KJ, Greenland S, editores. *Modern epidemiology*. 2nd ed. Boston: Lippincott Raven; 1998. p. 401–32.
96. Thompson WD. Statistical analysis of case-control studies. *Epidemiol Rev*. 1994;16:33–50.
97. Schlesselman JJ. *Case-control studies design, conduct, analysis*. NY, Oxford: Oxford University Press; 1982. p. 235–41.
98. Clayton D, Hills M. *Statistical models in epidemiology*. Oxford: Oxford University Press; 1993. p. 271–81.
99. Altman DG, Gore SM, Gardner MJ, et al. Statistical guidelines for contributors to medical journals. *BMJ*. 1983;286:1489–93.
100. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. (Versión electrónica actualizada en febrero de 2006. Disponible en: <http://www.icmje.org/>) *N Engl J Med*. 1997;336:309–15.
101. Mullner M, Matthews H, Altman DG. Reporting on statistical methods to adjust for confounding: a cross-sectional survey. *Ann Intern Med*. 2002;136:122–6.
102. Olsen J, Basso O. Residual confounding. *Am J Epidemiol*. 1999;149:290.
103. Hallan S, de Mutsert R, Carlsen S, et al. Obesity, smoking, and physical inactivity as risk factors for CKD: are men more vulnerable? *Am J Kidney Dis*. 2006;47:396–405.
104. Gotsche PC. Believability of relative risks and odds ratios in abstracts: cross sectional study. *BMJ*. 2006;333:231–4.
105. Szklo MF, Nieto J. *Epidemiology, beyond the basics*. Sudbury (MA): Jones and Bartlett; 2000. p. 408–30.
106. Chandola T, Brunner E, Marmot M. Chronic stress at work and the metabolic syndrome: prospective study. *BMJ*. 2006;332:521–5.
107. Vach W, Blettner M. Biased estimation of the odds ratio in case-control studies due to the use of ad hoc methods of correcting for missing values for confounding variables. *Am J Epidemiol*. 1991;134:895–907.
108. Little RJ, Rubin DB. *Statistical analysis with missing data*. NY: Wiley; 2002. p. 19–23.
109. Ware JH. Interpreting incomplete data in studies of diet and weight loss. *N Engl J Med*. 2003;348:2136–7.
110. Rubin DB. Inference and missing data. *Biometrika*. 1976;63:581–92.
111. Schafer JL. *Analysis of incomplete multivariate data*. London: Chapman & Hall; 1997.
112. Lipsitz SR, Ibrahim JG, Chen MH, et al. Non-ignorable missing covariates in generalized linear models. *Stat Med*. 1999;18:2435–48.
113. Rotnitzky A, Robins J. Analysis of semi-parametric regression models with non-ignorable non-response. *Stat Med*. 1997;16:81–102.
114. Rubin DB. *Multiple imputation for nonresponse in surveys*. NY: John Wiley; 1987.
115. Barnard J, Meng XL. Applications of multiple imputation in medical studies: from AIDS to NHANES. *Stat Methods Med Res*. 1999;8:17–36.
116. Braitstein P, Brinkhof MW, Dabis F, et al. Mortality of HIV-1-infected patients in the first year of antiretroviral therapy: comparison between low-income and high-income countries. *Lancet*. 2006;367:817–24.
117. Purandare N, Burns A, Daly KJ, et al. Cerebral emboli as a potential cause of Alzheimer's disease and vascular dementia: case-control study. *BMJ*. 2006;332:1119–24.
118. Steyn K, Gaziano TA, Bradshaw D, et al. Hypertension in South African adults: results from the Demographic and Health Survey, 1998. *J Hypertens*. 2001;19:1717–25.
119. Lohr SL. *Sampling: design and analysis*. Pacific Grove (CA): Duxbury Press; 1999. cap. 7.5.
120. Dunn NR, Arscott A, Thorogood M. The relationship between use of oral contraceptives and myocardial infarction in young women with fatal outcome, compared to those who survive: results from the MICA case-control study. *Contraception*. 2001;63:65–9.
121. Rothman KJ, Greenland S. Basic methods for sensitivity analysis and external adjustment. En: Rothman KJ, Greenland S, editores. *Modern epidemiology*. 2nd ed. Boston: Lippincott Raven; 1998. p. 343–57.
122. Custer B, Longstreth Jr WT, Phillips LE, et al. Hormonal exposures and the risk of intracranial meningioma in women: a population-based case-control study. *BMC Cancer*. 2006;6:152.
123. Wakefield MA, Chaloupka FJ, Kaufman NJ, et al. Effect of restrictions on smoking at home, at school, and in public places on teenage smoking: cross sectional study. *BMJ*. 2000;321:333–7.
124. Greenland S. The impact of prior distributions for uncontrolled confounding and response bias: a case study of the relation of wire codes and magnetic fields to childhood leukemia. *J Am Stat Assoc*. 2003;98:47–54.
125. Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology*. 2003;14:451–8.
126. Phillips CV. Quantifying and reporting uncertainty from systematic errors. *Epidemiology*. 2003;14:459–66.
127. Cornfield J, Haenszel W, Hammond EC, et al. Smoking and lung cancer: recent evidence and a discussion of some questions. *J Natl Cancer Inst*. 1959;22:173–203.
128. Langholz B. Factors that explain the power line configuration wiring code-childhood leukemia association: what would they look like? *Bioelectromagnetics*. 2001;5:S19–31.
129. Eisner MD, Smith AK, Blanc PD. Bartenders' respiratory health after establishment of smoke-free bars and taverns. *JAMA*. 1998;280:1909–14.
130. Dunne MP, Martin NG, Bailey JM, et al. Participation bias in a sexuality survey: psychological and behavioural characteristics of responders and non-responders. *Int J Epidemiol*. 1997;26:844–54.
131. Schuz J, Kaatsch P, Kaletsch U, et al. Association of childhood cancer with factors related to pregnancy and birth. *Int J Epidemiol*. 1999;28:631–9.

132. Cnattingius S, Zack M, Ekblom A, et al. Prenatal and neonatal risk factors for childhood myeloid leukemia. *Cancer Epidemiol Biomarkers Prev.* 1995;4:441–5.
133. Schuz J. Non-response bias as a likely cause of the association between young maternal age at the time of delivery and the risk of cancer in the offspring. *Paediatr Perinat Epidemiol.* 2003;17:106–12.
134. Slattery ML, Edwards SL, Caan BJ, et al. Response rates among control subjects in case-control studies. *Ann Epidemiol.* 1995;5:245–9.
135. Schulz KF, Grimes DA. Case-control studies: research in reverse. *Lancet.* 2002;359:431–4.
136. Olson SH, Voigt LF, Begg CB, et al. Reporting participation in case-control studies. *Epidemiology.* 2002;13:123–6.
137. Morton LM, Cahill J, Hartge P. Reporting participation in epidemiologic studies: a survey of practice. *Am J Epidemiol.* 2006;163:197–203.
138. Olson SH. Reported participation in case-control studies: changes over time. *Am J Epidemiol.* 2001;154:574–81.
139. Sandler DP. On revealing what we'd rather hide: the problem of describing study participation. *Epidemiology.* 2002;13:117.
140. Hepworth SJ, Schoemaker MJ, Muir KR, et al. Mobile phone use and risk of glioma in adults: case-control study. *BMJ.* 2006;332:883–7.
141. Hay AD, Wilson A, Fahey T, et al. The duration of acute cough in pre-school children presenting to primary care: a prospective cohort study. *Fam Pract.* 2003;20:696–705.
142. Egger M, Juni P, Bartlett C. Value of flow diagrams in reports of randomized controlled trials. *JAMA.* 2001;285:1996–9.
143. Osella AR, Misciagna G, Guerra VM, The Association for the Study of Liver Disease in Puglia. Hepatitis C virus (HCV) infection and liver-related mortality: a population-based cohort study in southern Italy. *Int J Epidemiol.* 2000;29:922–7.
144. Dales LG, Ury HK. An improper use of statistical significance testing in studying covariables. *Int J Epidemiol.* 1978;7:373–5.
145. Maldonado G, Greenland S. Simulation study of confounders election strategies. *Am J Epidemiol.* 1993;138:923–36.
146. Tanis BC, van den Bosch MA, Kemmeren JM, et al. Oral contraceptives and the risk of myocardial infarction. *N Engl J Med.* 2001;345:1787–93.
147. Rothman KJ, Greenland S, editors. *Modern epidemiology*. 2nd ed. Boston: Lippincott Raven; 1998. p. 120–5.
148. Clark TG, Altman DG, De Stavola BL. Quantification of the completeness of follow-up. *Lancet.* 2002;359:1309–10.
149. Qiu C, Fratiglioni L, Karp A, et al. Occupational exposure to electromagnetic fields and risk of Alzheimer's disease. *Epidemiology.* 2004;15:687–94.
150. Kengeya-Kayondo JF, Kamali A, Nunn AJ, et al. Incidence of HIV-1 infection in adults and sociodemographic characteristics of seroconverters in a rural population in Uganda: 1990–1994. *Int J Epidemiol.* 1996;25:1077–82.
151. Mastrangelo G, Fedeli U, Fadda E, et al. Increased risk of hepatocellular carcinoma and liver cirrhosis in vinyl chloride workers: synergistic effect of occupational exposure with alcohol intake. *Environ Health Perspect.* 2004;112:1188–92.
152. Salo PM, Arbes Jr SJ, Sever M, et al. Exposure to *Alternaria alternata* in US homes is associated with asthma symptoms. *J Allergy Clin Immunol.* 2006;118:892–8.
153. Pocock SJ, Clayton TC, Altman DG. Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. *Lancet.* 2002;359:1686–9.
154. Sasieni P. A note on the presentation of matched case-control data. *Stat Med.* 1992;11:617–20.
155. Lee GM, Neutra RR, Hristova L, et al. A nested case-control study of residential and personal magnetic field measures and miscarriages. *Epidemiology.* 2002;13:21–31.
156. Tiihonen J, Wallhbeck K, Lonnqvist J, et al. Effectiveness of antipsychotic treatments in a nationwide cohort of patients in community care after first hospitalisation due to schizophrenia and schizoaffective disorder: observational follow-up study. *BMJ.* 2006;333:224.
157. Christenfeld NJ, Sloan RP, Carroll D, et al. Risk factors, confounding, and the illusion of statistical control. *Psychosom Med.* 2004;66:868–75.
158. Smith GD, Phillips A. Declaring independence: why we should be cautious. *J Epidemiol Commun Health.* 1990;44:257–8.
159. Greenland S, Neutra R. Control of confounding in the assessment of medical technology. *Int J Epidemiol.* 1980;9:361–7.
160. Robins JM. Data, design, and background knowledge in etiologic inference. *Epidemiology.* 2001;12:313–20.
161. Sagiv SK, Tolbert PE, Altshul LM, et al. Organochlorine exposures during pregnancy and infant size at birth. *Epidemiology.* 2007;18:120–9.
162. World Health Organization. Body Mass Index (BMI). Disponible en: http://www.euro.who.int/nutrition/20030507_1 (acceso el 10 de septiembre de 2007).
163. Beral V. Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet.* 2003;362:419–27.
164. Hill AB. The environment and disease: association or causation? *Proc R Soc Med.* 1965;58:295–300.
165. Vineis P. Causality in epidemiology. *Soz Praventivmed.* 2003;48:80–7.
166. Empana JP, Ducimetiere P, Arveiler D, The PRIME Study. Are the Framingham and PROCAM coronary heart disease risk functions applicable to different European populations? *Eur Heart J.* 2003;24:1903–11.
167. Tunstall-Pedoe H, Kuulasmaa K, Mahonen M, et al. Contribution of trends in survival and coronary-event rates to changes in coronary heart disease mortality: 10-year results from 37 WHO MONICA project populations. Monitoring trends and determinants in cardiovascular disease. *Lancet.* 1999;353:1547–57.
168. Cambien F, Chretien JM, Ducimetiere P, et al. Is the relationship between blood pressure and cardiovascular risk dependent on body mass index? *Am J Epidemiol.* 1985;122:434–42.
169. Hosmer DW, Taber S, Lemeshow S. The importance of assessing the fit of logistic regression models: a case study. *Am J Public Health.* 1991;81:1630–5.
170. Tibshirani R. A plain man's guide to the proportional hazards model. *Clin Invest Med.* 1982;5:63–8.
171. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. *Am J Public Health.* 1998;88:15–9.
172. Uter W, Pfahlberg A. The application of methods to quantify attributable risk in medical practice. *Stat Methods Med Res.* 2001;10:231–7.
173. Schwartz LM, Woloshin S, Dvorin EL, et al. Ratio measures in leading medical journals: structured review of accessibility of underlying absolute risks. *BMJ.* 2006;333:1248.
174. Nakayama T, Zaman MM, Tanaka H. Reporting of attributable and relative risks, 1966–97. *Lancet.* 1998;351:1179.
175. Cornfield J. A method of estimating comparative rates from clinical data; applications to cancer of the lung, breast, and cervix. *J Natl Cancer Inst.* 1951;11:1269–75.
176. Pearce N. What does the odds ratio estimate in a case-control study? *Int J Epidemiol.* 1993;22:1189–92.
177. Rothman KJ, Greenland S. Measures of disease frequency. En: Rothman KJ, Greenland S, editores. *Modern epidemiology*. 2nd ed. Boston: Lippincott Raven; 1998. p. 44–5.
178. Doll R, Hill AB. The mortality of doctors in relation to their smoking habits: a preliminary report. 1954. *BMJ.* 2004;328:1529–33.
179. Ezzati M, López AD. Estimates of global mortality attributable to smoking in 2000. *Lancet.* 2003;362:847–52.
180. Greenland S. Applications of stratified analysis methods. En: Rothman KJ, Greenland S, editores. *Modern epidemiology*. 2nd ed. Boston: Lippincott Raven; 1998. p. 295–7.
181. Rose G. Sick individuals and sick populations. *Int J Epidemiol.* 2001;30:427–32, discussion 433–4.
182. Vandenbroucke JP, Koster T, Briet E, et al. Increased risk of venous thrombosis in oral-contraceptive users who are carriers of factor V Leiden mutation. *Lancet.* 1994;344:1453–7.
183. Botto LD, Khoury MJ. Commentary: facing the challenge of gene environment interaction: the two-by-four table and beyond. *Am J Epidemiol.* 2001;153:1016–20.
184. Wei L, MacDonald TM, Walker BR. Taking glucocorticoids by prescription is associated with subsequent cardiovascular disease. *Ann Intern Med.* 2004;141:764–70.
185. Martinelli I, Taioli E, Battaglioli T, et al. Risk of venous thromboembolism after air travel: interaction with thrombophilia and oral contraceptives. *Arch Intern Med.* 2003;163:2771–4.
186. Kyzas PA, Loizou KT, Ioannidis JP. Selective reporting biases in cancer prognostic factor studies. *J Natl Cancer Inst.* 2005;97:1043–55.
187. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. *Am J Epidemiol.* 1980;112:467–70.
188. Saracci R. Interaction and synergism. *Am J Epidemiol.* 1980;112:465–6.
189. Rothman KJ. *Epidemiology. An introduction*. Oxford: Oxford University Press; 2002. p. 168–80.
190. Rothman KJ. *Interactions between causes*. Modern epidemiology. Boston: Little Brown; 1986. p. 311–26.
191. Hess DR. How to write an effective discussion. *Respir Care.* 2004;49:1238–41.
192. Horton R. The hidden research paper. *JAMA.* 2002;287:2775–8.
193. Horton R. The rhetoric of research. *BMJ.* 1995;310:985–7.
194. Docherty M, Smith R. The case for structuring the discussion of scientific papers. *BMJ.* 1999;318:1224–5.
195. Perneger TV, Hudelson PM. Writing a research article: advice to beginners. *Int J Qual Health Care.* 2004;16:191–2.
196. *Annals of Internal Medicine*. Information for authors. Disponible en: http://www.annals.org/shared/author_info.html (acceso el 10 de septiembre de 2007).
197. Maldonado G, Poole C. More research is needed. *Ann Epidemiol.* 1999;9:17–8.
198. Phillips CV. The economics of “more research is needed”. *Int J Epidemiol.* 2001;30:771–6.
199. Winkleby MA, Kraemer HC, Ahn DK, et al. Ethnic and socioeconomic differences in cardiovascular disease risk factors: findings for women from the Third National Health and Nutrition Examination Survey, 1988–1994. *JAMA.* 1998;280:356–62.
200. Galuska DA, Will JC, Serdula MK, et al. Are health care professionals advising obese patients to lose weight? *JAMA.* 1999;282:1576–8.
201. Spearman C. The proof and measurement of association between two things. *Am J Psychol.* 1904;15:72–101.
202. Fuller WA, Hidioglou MA. Regression estimates after correcting for attenuation. *J Am Stat Assoc.* 1978;73:99–104.
203. MacMahon S, Peto R, Cutler J, et al. Blood pressure, stroke, and coronary heart disease. Part 1. Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet.* 1990;335:765–74.

204. Phillips AN, Smith GD. How independent are “independent” effects? Relative risk estimation when correlated exposures are measured imprecisely. *J Clin Epidemiol.* 1991;44:1223–31.
205. Phillips AN, Smith GD. Bias in relative odds estimation owing to imprecise measurement of correlated exposures. *Stat Med.* 1992;11:953–61.
206. Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol.* 1980;112:564–9.
207. Poole C, Peters U, Ilyasova D, et al. Commentary: this study failed? *Int J Epidemiol.* 2003;32:534–5.
208. Kaufman JS, Cooper RS, McGee DL. Socioeconomic status and health in blacks and whites: the problem of residual confounding and the resiliency of race. *Epidemiology.* 1997;8:621–8.
209. Greenland S. Randomization, statistics, and causal inference. *Epidemiology.* 1990;1:421–9.
210. Taubes G. Epidemiology faces its limits. *Science.* 1995;269:164–9.
211. Temple R. Meta-analysis and epidemiologic studies in drug development and postmarketing surveillance. *JAMA.* 1999;281:841–4.
212. Greenberg RS, Shuster Jr JL. Epidemiology of cancer in children. *Epidemiol Rev.* 1985;7:22–48.
213. Kushi LH, Mink PJ, Folsom AR, et al. Prospective study of diet and ovarian cancer. *Am J Epidemiol.* 1999;149:21–31.
214. Kemmeren JM, Algra A, Meijers JC, et al. Effect of second- and third-generation oral contraceptives on the protein C system in the absence or presence of the factor V Leiden mutation: a randomized trial. *Blood.* 2004;103:927–33.
215. Egger M, May M, Chene G, et al. Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. *Lancet.* 2002;360:119–29.
216. Campbell DT. Factors relevant to the validity of experiments in social settings. *Psychol Bull.* 1957;54:297–312.
217. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med.* 1999;130:515–24.
218. Krinsky S, Rothenberg LS, editors. Conflict of interest policies in science and medical journals: editorial practices and author disclosures. *Sci Eng Ethics.* 2001;7:205–18.
219. Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *JAMA.* 2003;289:454–65.
220. Davidson RA. Source of funding and outcome of clinical trials. *J Gen Intern Med.* 1986;1:155–8.
221. Stelfox HT, Chua G, O'Rourke K, et al. Conflict of interest in the debate over calcium-channel antagonists. *N Engl J Med.* 1998;338:101–6.
222. Lexchin J, Bero LA, Djulbegovic B, et al. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ.* 2003;326:1167–70.
223. Als-Nielsen B, Chen W, Gluud C, et al. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA.* 2003;290:921–8.
224. Barnes DE, Bero LA. Why review articles on the health effects of passive smoking reach different conclusions? *JAMA.* 1998;279:1566–70.
225. Barnes DE, Bero LA. Industry-funded research and conflict of interest: an analysis of research sponsored by the tobacco industry through the Center for Indoor Air Research. *J Health Polit Policy Law.* 1996;21:515–42.
226. Glantz SA, Barnes DE, Bero L, et al. Looking through a keyhole at the tobacco industry. The Brown and Williamson documents. *JAMA.* 1995;274:219–24.
227. Huss A, Egger M, Hug K, et al. Source of funding and results of studies of health effects of mobile phone use: systematic review of experimental studies. *Environ Health Perspect.* 2007;115:1–4.
228. Safer DJ. Design and reporting modifications in industry sponsored comparative psychopharmacology trials. *J Nerv Ment Dis.* 2002;190:583–92.
229. Aspinall RL, Goodman NW. Denial of effective treatment and poor quality of clinical information in placebo controlled trials of ondansetron for post-operative nausea and vomiting: a review of published trials. *BMJ.* 1995;311:844–6.
230. Chan AW, Hrobjartsson A, Haahr MT, et al. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA.* 2004;291:2457–65.
231. Melander H, Ahlqvist-Rastad J, Meijer G, et al. Evidence b(i)ased medicine—selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ.* 2003;326:1171–3.
232. Scherer RW, Langenberg P, von Elm E. Publicación de resultados inicialmente presentada en resumen. *Cochrane Database of Systematic Reviews*, Tomo 2. Art. No.: MR000005. Disponible en: <http://www.cochrane.org/reviews/en/mr000005.html> (acceso el 10 de septiembre de 2007).
233. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet.* 2001;357:1191–4.
234. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of observational studies in epidemiology. *JAMA.* 2000;283:2008–12.
235. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med.* 2001;134:663–94.
236. Moher D. CONSORT: an evolving tool to help improve the quality of reports of randomized controlled trials. Consolidated standards of reporting trials. *JAMA.* 1998;279:1489–91.
237. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA.* 1996;276:637–9.