
Discordancias entre los estudios de ámbitos hospitalario y comunitario cuando evalúan la misma pregunta de investigación

M. Delgado Rodríguez
Universidad de Jaén

Correspondencia: Dr. Miguel Delgado Rodríguez. Cátedra de Medicina Preventiva. Universidad de Jaén. Edificio B-3. 23071 Jaén. Correo electrónico: mdelgado@ujaen.es

Recibido: 25 de septiembre de 2001.

Aceptado: 15 de febrero de 2002.

(Disagreement among hospital and community studies evaluating the same research question)

Resumen

Se abordan las razones que motivan que los resultados de la investigación realizada sobre poblaciones hospitalarias en ocasiones no sea consistente con la practicada en la comunidad. En la base de las diferencias se encuentran el proceso de selección de los pacientes y la recogida de datos por la historia clínica (su falta de uniformidad y el tratamiento de los «no consta»). Se estructura la razón de las discrepancias en función del tipo de pregunta de investigación: frecuencia, diagnóstico, etiología, pronóstico y tratamiento-prevención. Es el proceso de selección el que justifica las diferencias en frecuencia y pronóstico. En relación con el diagnóstico, las discrepancias se deben a la prevalencia de la enfermedad. En el estudio de asociaciones de causalidad ciertos errores son más frecuentes en los estudios hospitalarios, como los sesgos de detección, protopático (ambos exageran la asociación), y el de inclusión (reduce la asociación). Se analizan ejemplos publicados de cada una de ellas. En relación con el tratamiento-prevención, los problemas son sobre todo de validez externa, ya que la metodología del ensayo clínico previene las amenazas a la validez interna; se analiza con un ejemplo de valoración de la eficacia vacunal en pacientes y población sana. La frecuencia de citación de los sesgos se midió en una búsqueda en Medline; en los estudios hospitalarios se mencionó con más frecuencia el sesgo de detección ([RR] = 2,71; intervalo de confianza [IC] del 95%, 1,69-4,37) y el sesgo de confusión por indicación (RR = 1,76; IC del 95%, 0,90-3,42). Por último, se da una serie de recomendaciones destinadas a aumentar la validez de los estudios realizados en el medio hospitalario.

Palabras clave: Estudios hospitalarios. Estudios comunitarios. Sesgos. Concordancia.

Abstract

The goal of this review is to delineate some of reasons that justify the lack of consistency between hospital-based and community research. The main reasons for the differences are the selection of the hospital population and information based on clinical chart (its lack of uniformity and the treatment of the «not available» data). The reasons for lack of consistency are divided according to the type of research question: frequency, diagnosis, etiology, prognosis and treatment-prevention. The way a hospital population is selected justifies discrepancies regarding frequency and prognosis. As regards diagnosis, differences are mainly due the prevalence of disease. In the ascertainment of causality several biases are more common in hospital-based research, such as detection bias, protopathic (both producing an away-from-null estimate), and inclusion bias (diminishing the strength of association). Examples taken from the medical literature are offered to illustrate each bias. Regarding treatment-prevention problems arise from external validity, as clinical trials are less prone to bias; this latter situation is exemplified with an assessment of vaccine efficacy in both patients and healthy population. The frequency of citation of bias was assessed by a Medline search; in hospital studies detection bias and confounding by indication were more often quoted than in non-hospital research (RR = 2.71; 95% CI; 1.69-4.37; RR = 1.76; 95% CI, 0.90-3.42, respectively). Lastly, several recommendations are given to increase the validity of hospital-based research.

Key words: Hospital studies. Community studies. Bias. Agreement.

Introducción

Por el marcado carácter inductista de la investigación no basta con la realización de un estudio para responder a una pregunta de investigación, sino que para valorar cualquier asociación se practican varios, y con frecuencia en diferentes ámbitos, el hospital y la comunidad. Con una frecuencia no determinada los resultados de las investigaciones realizadas en el hospital no coinciden (no son consistentes según el tercer principio de Hill¹). No hemos localizado ninguna revisión en la que se analicen y se ilustren mediante ejemplos reales las razones por las que los resultados de los estudios hospitalarios y comunitarios referidos al mismo tema difieran.

En 1946 Joseph Berkson ponía de manifiesto por vez primera que los resultados que se observaban en un estudio realizado en un hospital no se correspondían con lo que sucedía en la población de referencia²; daba origen así al sesgo que lleva su nombre y que se ha tratado de manera detallada en la literatura científica³⁻⁶. El sesgo de Berkson era la primera llamada de atención a que la investigación hospitalaria no tenía por qué ser coincidente con la investigación comunitaria. Con posterioridad se han incorporado otras fuentes de error, que serán objeto de la presente revisión. A pesar de las dificultades metodológicas, el hospital sigue siendo una fuente muy importante de investigación epidemiológica, ya que supone una concentración de los casos de enfermedad que se generan en la comunidad.

El objetivo de la presente revisión es ilustrar mediante ejemplos las razones que pueden motivar la existencia de discrepancia (o falta de consistencia) entre los estudios, realizados en el hospital y en la comunidad, que intentan responder a la misma pregunta de investigación. Para ello se indicarán en primer lugar las particularidades básicas de la investigación hospitalaria que pueden suponer diferencias de partida con respecto a la investigación comunitaria, a continuación se enunciarán las discrepancias en función de la pregunta de

investigación, se intentará establecer la frecuencia de las discordancias y se finalizará con una serie de recomendaciones para la realización de una investigación hospitalaria de calidad.

Particularidades de la investigación hospitalaria

Radican esencialmente en dos apartados: en el proceso de selección de la población de estudio y en la recogida de datos, que se basa especialmente en la historia clínica.

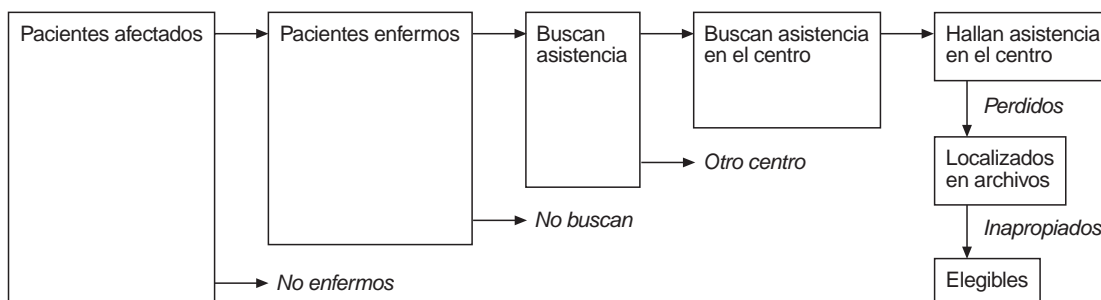
Proceso de selección

Sea cual fuere la línea de trabajo, los investigadores deben ser conscientes del hecho básico de que recabar información sobre pacientes supone tener en cuenta el proceso por el que un paciente acude al hospital, que se resume en la figura 1 (modificada de Delgado Rodríguez⁷). La conclusión que se debe extraer de esta figura es que los elegidos no tienen por qué representar a los afectados (se puede producir un sesgo de selección) y proporciona un esquema básico que se utilizará con posterioridad en la presente revisión.

Recogida de datos: la historia clínica

En las instituciones sanitarias, la historia clínica es fuente de numerosos trabajos de investigación, aunque en general los profesionales sanitarios reconocen que la mayor parte de las historias clínicas tienen un fin asistencial, por lo que no cumplen todas las exigencias de una buena recogida de información. El mayor inconveniente que presenta el uso de la historia clínica es el problema de los datos que no constan. Un dato que no consta puede deberse a varios hechos: que no se

Figura 1. Selección de la población de estudio en un hospital.



haya valorado, que siendo medido ha dado un resultado negativo, o incluso que aunque haya dado un resultado positivo no se encuentre reflejado en la historia. No se puede saber cuál ha sido el caso y los resultados pueden cambiar en función de las asunciones que se hagan. Por ejemplo, Horwitz y Stewart⁸ realizaron un estudio de casos y controles hospitalarios sobre cáncer de mama y terapia estrogénica con recogida de datos de la historia clínica. Cuando se eliminaron los «no consta» de la información acerca del tratamiento con estrógenos, la *odds ratio* [OR] fue de 0,9, indicativa de ausencia de relación; por contra, cuando se asumió que los «no consta» eran que no se había prescrito la terapia hormonal sustitutiva, la OR fue sensiblemente distinta, de 3,3 (IC del 95%, 2,2-5,0). Lo anterior refleja el carácter interpretativo de la información contenida en la historia clínica. Por tanto, la ausencia de ciertos datos no implica la no presencia de una cierta variable, sino que no se consideró en su momento relevante en el contexto de la situación del paciente. Esto se ha valorado en el estudio de factores pronósticos de mortalidad en la cirugía coronaria, comparando con la información extraída tras la lectura de toda la historia clínica⁹. Las diferencias fueron muy importantes; en general hubo una tendencia a la infrainformación de las comorbilidades. Esto se esperaría que fuera menor en los fallecidos (justificar mejor las razones de la muerte), con lo que el sesgo introducido sería positivo (un riesgo relativo por encima del real). Sin embargo, no se encontró uniformidad en este aspecto.

Los estudios hospitalarios pueden tener una tendencia a utilizar más los datos de la historia, mientras que los estudios comunitarios necesariamente han de utilizar datos de entrevista. Como se ha argumentado los resultados no tienen por qué ser coincidentes. Un ejemplo de esta situación se hizo para estudiar el alcohol como factor de riesgo del bajo peso al nacimiento: los datos de la historia dieron un valor de OR contrario al que se obtuvo con los datos de la entrevista¹⁰.

El otro gran problema que presenta la historia clínica es la falta de uniformidad (estandarización) en las pruebas realizadas o en la anotación de los resultados. Por ejemplo, el consumo de alcohol o tabaco puede recogerse con palabras («moderado», «intenso», etc.), con flechas (↑↑↑, ↑↑, etc.) o cruces (+, ++, etc.). ¿Cuál es la equivalencia entre estos términos? A ello habría que añadir que la anotación puede estar influida por la experiencia u opiniones del anotador. Un no bebedor puede reflejar el hábito alcohólico de un paciente de manera distinta a como lo haría un consumidor habitual de alcohol. El uso de la información clínica no contrastada puede introducir errores importantes, especialmente para cierto tipo de variables.

En la historia (como en cualquier otro registro) se anotan los resultados numéricos de las pruebas diagnósticas y terapéuticas realizadas en los pacientes. Los

datos numéricos en ocasiones no se reflejan con precisión total, sino de manera aproximada y los anotadores pueden variar en la preferencia de la cifra final a la que redondean: los hay que prefieren los números pares a los impares, los que redondean por abajo o por arriba para acabar en 0 (un colesterol de 244 mg puede reflejarse como 240 o 250), etc. Como es natural esto introduce un error¹¹.

Discrepancias según la línea de investigación

Las preguntas de investigación se pueden agrupar de una manera sencilla, simplificando las ideas de Hullely y Cummins¹², dentro de cualquiera de las cinco líneas siguientes: frecuencia, diagnóstico, etiología, pronóstico y tratamiento-prevención.

Estimación de la frecuencia de un proceso

Los centros asistenciales concentran enfermos y los profesionales se ven tentados con frecuencia para utilizar estos datos y proporcionar así una aproximación de la magnitud de un proceso. Esto sería sólo válido cuando todas las disyuntivas sistematizadas en la figura 1 no existen y los elegidos son igual a los afectados. Dentro del esquema de esta figura cobran un interés especial los procesos en los que no todos los afectados se sienten enfermos y cuando no todos los afectados buscan asistencia. Schoenberg et al¹³ realizaron un estudio de prevalencia de la enfermedad de Parkinson en el condado de Copiah (Mississippi, EE.UU.) en el que se abordó mediante una encuesta a todo el condado, puerta por puerta. A los que declararon síntomas sugestivos de la enfermedad se les remitió para un examen neurológico completo. Tras la detección en dos etapas se encontró un total de 31 casos en los casi 24.000 habitantes, de los que 13 no habían ido a ningún centro de asistencia sanitaria y habían pasado completamente inadvertidos, las razones de ello es que las manifestaciones de la enfermedad se atribuían simplemente a la vejez o no impedían otras funciones del individuo.

También debe cuestionarse la frecuencia de un proceso que se estima con los datos de centros de referencia, según el esquema de la figura 1.

Valoración del diagnóstico

Por regla general, la mayoría de las investigaciones en esta línea se centra en instituciones asistenciales. La excepción es la utilización de una prueba con fines de cribado. No se ilustrará este apartado porque las ra-

zonas se conocen desde hace tiempo y no merecen un análisis detallado; se deben sobre todo al espectro de la enfermedad –en los casos avanzados aumenta la sensibilidad–, al grupo de referencia –que altera la especificidad¹⁴– y a la prevalencia, que influye sobre todos los parámetros de una prueba, aunque notablemente más sobre los valores predictivos¹⁵.

Valoración de la etiología

Aquí, en general, se puede producir una serie de errores. No se comentará el sesgo de Berkson, porque ha recibido un tratamiento extenso en la literatura científica²⁻⁶, pero sí otras situaciones en las que se pueden presentar otros errores.

La frecuencia de la exposición influye en el valor de la fuerza de la asociación. En el estudio de la etiología o relación de causalidad no se puede olvidar que la estimación de la frecuencia de la exposición ha de ser válida, ya que según el modelo de causalidad de Rothman¹⁶ ésta influye en el valor de la fuerza de la asociación. Este parámetro es fácil que cambie según el grado de referencia que ocupa un centro. También es posible que se produzca si la exposición es una intervención sanitaria, que puede cambiar según el grado de aplicación en ciertas instituciones o sectores de la población. Morgenstern et al¹⁷ ofrecen un ejemplo de esta última situación cuando analizaron la relación entre los neurolépticos y la discinesia tardía mediante un metaanálisis de 21 estudios con seguimiento, 14 de ellos en el ámbito hospitalario y siete en otros ámbitos sanitarios (seis en asilos y uno en pacientes psiquiátricos ambulatorios). Las estimaciones de OR derivadas de los estudios hospitalarios y los realizados en otras poblaciones fueron muy diferentes: 5,3 frente a 2,6. Los autores aproximaron ambas estimaciones cuando realizaron un ajuste en función del tamaño del grupo expuesto (estandarización interna): 2,9 frente a 2,7. La única razón verosímil para justificar esto es la deducción del modelo de causalidad del Rothman mencionada al principio del párrafo.

El sesgo de detección. El sesgo de detección se produce cuando el efecto se detecta más en el grupo de los expuestos que en el de los no expuestos y su consecuencia es una sobrestimación de la fuerza de asociación. *A priori* el sesgo de detección debe ser un error al que deben ser más proclives las investigaciones realizadas en el hospital, porque tienen todas las herramientas diagnósticas más accesibles y sofisticadas. No es difícil encontrar en la bibliografía científica valoraciones de este error. A continuación se exponen dos ejemplos. El primero es el elegante metaanálisis de Bernal et al¹⁸ sobre la asociación entre vasectomía y cáncer de próstata. Cuando se combinaron cinco estudios

de cohortes y nueve de casos y controles resultó una asociación estadísticamente significativa por el procedimiento de DerSimonian-Laird (RR = 1,23; IC del 95%, 1,01-1,49). En su análisis de la heterogeneidad (fig. 2) observaron que los estudios comunitarios daban estimaciones más bajas que los hospitalarios (1,12 [IC del 95%, 0,96-1,32] frente a 1,98 [IC del 95%, 1,37-2,86]). Obsérvese que los intervalos de confianza no se solapan, lo que indica la existencia de diferencias significativas entre ambas estimaciones. En un análisis posterior valoran el tipo de diseño, el proceso de selección y la posible existencia de sesgo de detección (particularmente importante en esta asociación). Los estudios de casos y controles de ámbito hospitalario fueron los que dieron los valores más elevados de riesgo relativo, e igualmente sucedió con los estudios en los que la selección se consideró inadecuada y con probable sesgo de detección (mayoritariamente casos y controles hospitalarios). Los autores concluyeron que no había asociación y utilizaron como patrón la investigación comunitaria.

El segundo ejemplo es sobre la asociación entre la colestectomía y el riesgo ulterior de cáncer colorrectal. En efecto, el cáncer es una posible complicación derivada de una intervención, la colestectomía, que se realiza en el hospital y en donde, *a priori*, el hospital debe constituir la base del estudio. Giovannucci et al¹⁹ estudiaron esta relación mediante un metaanálisis en el que se combinaron las evidencias de cinco estudios de cohortes y 28 de casos y controles (ocho de ellos de ámbito poblacional). Globalmente se apreció una gran heterogeneidad (fig. 3) en la estimación ponderada global. Los estudios de cohortes dieron una estimación de RR inferior a la de los de casos y controles (0,97 [IC del 95%, 0,82-1,14] frente a 1,34 [IC del 95%, 1,14-1,57]). Cuando los estudios de casos y controles se dividieron en comunitarios y hospitalarios los

Figura 2. Vasectomía y cáncer de próstata: análisis de la heterogeneidad. (Fuente: Bernal-Delgado et al¹⁸.)

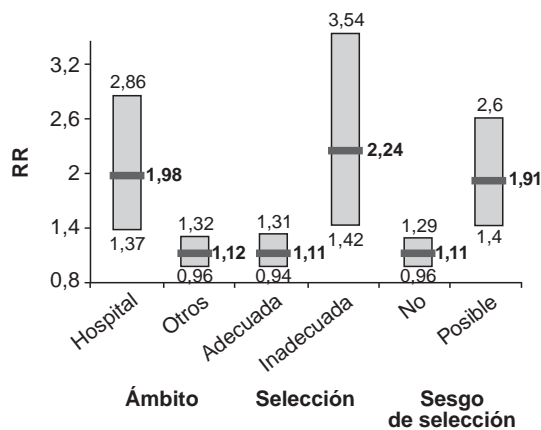
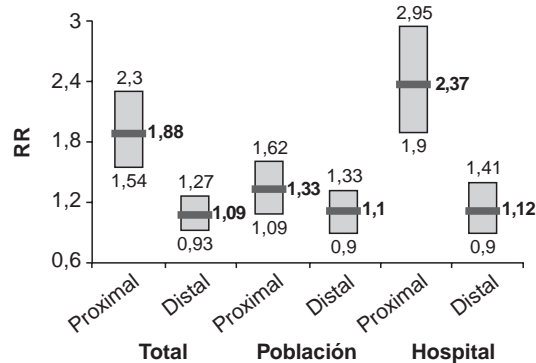


Figura 3. Colectomía y cáncer gástrico: análisis de la heterogeneidad. (Fuente: Giovannucci et al¹⁹.)



resultados fueron distintos, y los hospitalarios evidenciaron un RR más elevado que el de los comunitarios, que se aproximaron a los estudios de cohortes (1,17 [IC del 95%, 1,01-1,37]) frente a 1,43 [IC del 95%, 1,12-1,81]). En un intento de justificar los resultados, se estratificaron los estudios comunitarios y los hospitalarios en función del lugar anatómico de ocurrencia del cáncer: proximal y distal. Aquí es donde se observó la mayor discordancia entre los estudios de población y los hospitalarios, en el cáncer proximal el RR fue muy superior en los estudios hospitalarios (1,33 [IC del 95%, 1,09-1,62]) frente a 2,37 [IC del 95%, 1,90-2,95]); apréciase que los IC no se solapan, lo que indica la existencia de una diferencia estadísticamente significativa entre ambos resultados. No hubo diferencias en el cáncer de colon distal entre estudios hospitalarios y comunitarios (RR = 1,1-1,12). Los autores concluyeron que había una asociación, para la que no se ha documentado ninguna base biológica. No obstante, la vesícula biliar se encuentra al lado del colon ascendente o proximal. ¿No será que los estudios realizados en el hospital pueden sufrir un sesgo de detección, ya que los pacientes colecistectomizados son mejor estudiados y hay una mayor frecuencia de cánceres en el colon proximal, que son los más difíciles de detectar?

El sesgo de inclusión. El sesgo de inclusión se produce en los estudios de casos y controles cuando en el grupo de controles se incorporan individuos con procesos que mantienen relación con la exposición y produce una subestimación de la magnitud de la asociación. Al igual que el sesgo de detección, el sesgo de inclusión es más probable en la investigación hospitalaria que usa el estudio de casos y controles, algo muy conocido. Un ejemplo es la relación documentada entre *Helicobacter pylori* y el cáncer de estómago. La investigación de la exposición a *H. pylori* es más frecuente en el hospital, porque exige la realización de una prue-

ba de laboratorio (la detección de anticuerpos frente al microorganismo). Esta relación fue estudiada en un metaanálisis por Huang et al²⁰. Hasta ese momento se habían publicado 19 estudios sobre esa asociación, todos ellos de casos y controles, aunque con muy diferente procedencia de los controles. La OR ponderada global fue claramente significativa, 2,29 (IC del 95%, 2,04-2,58), no obstante, la heterogeneidad fue muy significativa. Uno de los factores clave que ayudaron a justificar el resultado fue el análisis de la procedencia de controles: cuando los controles procedieron de poblaciones comunitarias el riesgo fue sensiblemente mayor que cuando se extrajeron de poblaciones de pacientes del hospital (2,89 [IC del 95%, 2,51-3,43] frente a 1,37 [IC del 95%, 1,11-1,70]). El análisis de los grupos de los cuales se extrajeron estos controles plantea la posibilidad del sesgo de inclusión, la comunidad etiológica con los procesos que constituyen el grupo de referencia, que puede hacer más frecuente la seropositividad frente a *H. pylori*. De hecho, es llamativo en este metaanálisis que el estudio con controles comunitarios que encontró la OR más baja (un estudio portugués con una OR de 0,54, publicado en una revista no incluida en las revistas del Science Citation Index) emparejó a los controles por la existencia de afección gastrointestinal previa, que puede ser una etapa intermedia en la producción del efecto.

El sesgo protopático y confusión por indicación. Se confunden estos términos con frecuencia²¹. El sesgo protopático se produce cuando los estadios iniciales del efecto, normalmente subclínicos, condicionan un cambio en el nivel de exposición; tiene como consecuencia general producir una sobreestimación de la fuerza de la asociación. Este error se puede presentar en cualquier tipo de investigación, pero cuando la enfermedad esté más evolucionada, que es como se diagnostica en los hospitales, su probabilidad de presentación aumenta. Un ejemplo de ello lo constituye la asociación entre alcohol y litiasis biliar. Thijs et al²² encontraron que el consumo de dos o más bebidas alcohólicas daba una OR de 0,6, lo que sugería que el consumo de alcohol se comportaba como factor protector. Se les ocurrió pensar que era posible que las personas con litiasis biliar, antes de ser diagnosticadas del proceso, comenzasen con molestias digestivas que podrían verse agravadas por el consumo de alcohol. Así, los sujetos en la fase inicial de la enfermedad podrían tender a abandonar el consumo de alcohol, por las molestias producidas. Para comprobar esta hipótesis establecieron criterios de restricción en sus casos. En primer lugar, se centraron en los casos de enfermedad con comienzo agudo y sin molestias previas. En este tipo de pacientes, la OR no fue ya protectora, e incluso fue superior a la unidad (1,5). En segundo lugar, utilizaron a pacientes con litiasis biliar completamente asintomática que había sido

diagnosticada en un cribado o en un examen de salud. En este tipo de sujetos, la OR fue de 1. Los datos anteriores sugieren la presencia de un sesgo protopático en el estudio de la relación alcohol-litiasis biliar.

El sesgo por indicación (confusión por indicación) se presenta cuando se analizan medicaciones o intervenciones como consecuencia de procesos asociados con el efecto (la indicación de intervención es un auténtico factor de confusión), o como consecuencia de la gravedad del proceso. Esto último, por ejemplo, se ha sugerido para justificar la asociación entre la mortalidad por asma y fenoterol. Un estudio de seguimiento canadiense ha comprobado que el cambio de salbutamol a fenoterol se produce por un aumento de la gravedad del asma, aunque la elección inicial entre ambas medicaciones no se asoció a la gravedad²³. De esta manera, en los casos graves tratados más frecuentemente con fenoterol hay más mortalidad que en los casos leves, en los que abunda más el tratamiento con salbutamol.

En ocasiones es difícil separar el sesgo protopático del sesgo de indicación. Esto sucede, por ejemplo, en la asociación entre la cimetidina y el cáncer gástrico²⁴, en la que es probable que las manifestaciones iniciales del cáncer provoquen un mayor uso de la cimetidina y/o que otro proceso (la úlcera péptica) actúe como factor de confusión. En esta situación ayuda el análisis de latencia entre la intervención y el efecto, en el que se aprecia que típicamente desaparece conforme aumenta el tiempo de latencia entre el consumo y la aparición del efecto. Hay que garantizar, como siempre en todos los estudios, que la exposición/enfermedad no se valore de manera diferencial en los grupos que se comparan.

Este tipo de errores son también frecuentes en el entorno comunitario; por ejemplo, en el análisis de la fertilidad también hay que tenerlos en cuenta, ya que la mujer (más que el varón) que tiene problemas para quedar embarazada puede haber cambiado su conducta y sus hábitos de vida (disminuir el consumo de tabaco, café, etc.), y estos factores relacionarse posteriormente con problemas de fertilidad²⁵.

Valoración del pronóstico

No se estudiará el pronóstico que, aunque forma parte de la historia natural, tiene un marcado carácter hospitalario y las inconsistencias que se pueden encontrar entre los diferentes estudios radican, sobre todo, en el tipo de centro en el que se realizan; esto es, la posición de referencia que ocupan en un sistema de atención escalonado, el patrón de remisión al centro y la selección en función del estadio de la enfermedad²⁶, donde se altera el proceso de selección expuesto en la figura 1.

Valoración del tratamiento-prevención

La mayoría de los estudios de tratamiento tienen un ámbito institucional, mientras que en los de prevención suele ser comunitario, y entre ellos priman sobre todo los estudios experimentales. Las técnicas de aleatorización y enmascaramiento reducen en ellos la probabilidad de que exista un sesgo. No obstante, los problemas pueden venir en este caso por la validez externa (generalización a otras poblaciones). Esto se ilustra con el caso de la vacunación antineumocócica. La eficacia de esta vacunación se conoce ya desde algunos años y se han realizado varios metaanálisis. En uno de ellos se ofrecen los datos suficientes para tipificar la población en la que se probó la eficacia de la vacunación²⁷. Los autores no realizaron un análisis en función de la población de procedencia de los nueve ensayos clínicos que combinaron, pero sí ofrecieron un análisis que refleja bastante bien la heterogeneidad en función del tipo de población. Tan sólo elegiremos uno de los efectos que valoraron: la prevención de la neumonía neumocócica cierta (se valoró también la mortalidad, la bronquitis, otras neumonías y la posible neumonía neumocócica). Hubo una heterogeneidad altamente significativa en el estimador global de la eficacia, de tal manera que sacar la conclusión general de que la vacuna era eficaz era muy arriesgado (diferencia de riesgos [DR] = 0,004; IC del 95%, 0,000-0,007). Los autores estratificaron, por riesgo de neumonía de los pacientes, alto y bajo. Todos los pacientes de alto riesgo procedían del ámbito hospitalario (cinco estudios), mientras que los de bajo riesgo (cuatro estudios) eran comunitarios, a excepción en este último grupo de un estudio que incorporó a pacientes ambulatorios y pacientes hospitalizados psiquiátricos (en principio con procesos sistémicos menos graves). La vacuna fue eficaz en los sujetos de bajo riesgo (DR = 0,011; IC del 95%, 0,002-0,019), no en los de alto riesgo (DR = 0; IC del 95%, -0,001-0,002). Se ha de aclarar que en el estudio con pacientes ambulatorios y psiquiátricos que se consideraron de bajo riesgo la vacuna no fue eficaz, y sí lo fue en grado variable (pero en los tres de manera significativa) en los estudios comunitarios. Además, la heterogeneidad disminuyó considerablemente con esta estratificación. Éste es un ejemplo en el que la población ofrecida por el hospital no es la correcta, y aunque el ensayo clínico aleatorizado neutraliza los sesgos de selección que se pueden observar en otros tipos de investigaciones, hay un problema de validez externa.

Frecuencia de discrepancias entre la investigación hospitalaria y la comunitaria

Contestar a este apartado es difícil. La búsqueda de discordancias entre la investigación hospitalaria y

la comunitaria se ha realizado a través de una consulta en Medline. Dado que no hay un término MeSH para diferencias o discrepancias se han utilizado varios términos ingleses equivalentes sacados del Webster's New Dictionary of Synonyms (1984) en combinación con estudio hospitalario y comunitario con la siguiente estrategia: (*differen* or diverse or divergen* or disparate or inconsisten* or incoherenc* or incongruous* or discrepan* or discordan**) and (*community stud**) and (*hospital stud**). La búsqueda se realizó en todos los campos y para el período 1981-2001. Tan sólo se obtuvieron tres citas, de las que sólo una comparaba estudios comunitarios con estudios hospitalarios, referida al papel del alcohol en el cáncer de mama (Roth et al. J Clin Epidemiol 1994;47:207-16). Esto es un resultado claramente en contradicción con los ejemplos enumerados con anterioridad.

En una etapa posterior se valoró la frecuencia de mención de algunos de los errores comentados buscando de manera independiente los siguientes términos para el mismo período: *detection bias* (104 citas), *inclusion bias* (tres citas, de las que sólo una era relevante), y *protopathic bias or confounding by indication or indication bias* (56 citas). Como denominador para los sesgos se contabilizó en Medline entre 1981 y 2001 un total de 36.314 estudios hospitalarios de observación (5.704 de cohortes, 13.187 de casos y controles y 17.423 de corte) y 75.840 no hospitalarios (39.985 de cohortes, 12.408 de casos y controles y 23.447 de corte). De las 104 citas que mencionaban el sesgo de detección 39 pudieron atribuirse a estudios hospitalarios, 30 a estudios no hospitalarios y 35 no eran relevantes (19) o había dudas (16). Con los datos anteriores se estimó que el riesgo relativo entre estudio hospitalario y tratar el sesgo de detección fue de 2,71 (IC del 95%, 1,69-4,37) (tabla 1). Esto no significa que en el proceso de revisión sea más frecuente el sesgo de detección en los estudios hospitalarios, sino que lo mencionan más porque los autores conocen su amenaza.

Con respecto a las 56 citas que mencionaban el sesgo protopático o confusión por indicación, 19 eran

referentes a tratamientos administrados de manera ambulatoria, 16 a situaciones hospitalarias y 21 eran contribuciones en las que la referencia en Medline impedía la atribución o eran metodologías. La asociación entre estudio hospitalario y mención del sesgo no fue estadísticamente significativa (RR = 1,76, IC del 95%, 0,90-3,42).

Comentario

La valoración de errores presenta el problema de no tener un criterio de verdad, esto es, cuando se analizan las discrepancias entre estudios realizados en el ámbito hospitalario y comunitario surge la pregunta inevitable: ¿cuáles son los válidos? Se podría sacar la conclusión de que, acorde con los ejemplos ilustrados, ésta radica en los estudios comunitarios, no sometidos al proceso de selección particular que suele existir en los realizados en torno a instituciones. Sería arriesgado afirmar esto, en primer lugar, porque los ejemplos propuestos no provienen de una búsqueda sistemática. Es más, ha sido imposible sacar a la luz todas aquellas preguntas de investigación en las que, existiendo estudios realizados en diferentes ámbitos, éstos discrepan en sus conclusiones. Una posible fuente para conocerlas es el metaanálisis, pero los autores no reflejan, aún existiendo las diferencias, este aspecto en el resumen; además, no todas las preguntas en las que hay discrepancias han sido objeto de una revisión sistemática. Por tanto, enumerar una base poblacional de discrepancias es una tarea que desborda y que requeriría un saber enciclopédico de todo el ámbito biomédico, algo que creemos, si no imposible, francamente difícil.

En segundo lugar, no siempre se puede penalizar a los estudios hospitalarios frente a los comunitarios porque hay otras situaciones en las que los estudios hospitalarios no proporcionan la misma información que los asentados en la comunidad y no se encuentra una razón definida para estas diferencias. Es lo que se ha observado en varios metaanálisis, por ejemplo entre el cáncer invasivo de cérvix y la anticoncepción hormonal oral²⁸, entre la duración de la terapia estrogénica sustitutiva en la menopausia y el riesgo del cáncer de mama²⁹ o entre la misma exposición con el cáncer epitelial ovárico³⁰. En los trabajos anteriores no se puede identificar con claridad el origen de la diferencia, ya que los estudios incluidos en principio se ajustaban a una doctrina correcta de investigación.

Según lo anterior, ante la ausencia de un criterio de verdad concluyente sobre la validez de una investigación, en teoría no deberían minusvalorarse los estudios hospitalarios frente a los comunitarios. No obstante, en esta revisión se insiste en que con frecuencia la in-

Tabla 1. Citación de algunos sesgos y ámbito de la investigación encontrados en una búsqueda Medline en el período 1981-2001

Ámbito	N.º de citas	Mencionan sesgo (por 10.000)	RR (IC del 95%)
		Sesgo de detección	
Hospital	36.314	39 (10,7)	2,71 (1,69-4,37)
No hospital	75.840	30 (3,9)	1 (referencia)
		Sesgo protopático o confusión por indicación	
Hospital	36.314	16 (4,4)	1,76 (0,90-3,42)
No hospital	75.840	19 (2,5)	1 (referencia)

RR: riesgo relativo; IC: intervalo de confianza.

vestigación hospitalaria, por las razones metodológicas ilustradas, ofrece resultados inferiores en calidad a los producidos en otros ámbitos.

En la elección de los ejemplos se ha preferido la aproximación que ofrece el metaanálisis (estudios de población, ya sean cohortes o de casos y controles), frente a la búsqueda de estudios individuales que tengan más de un grupo de referencia (hospitalario y comunitario), con resultados discordantes³¹. Esta última estrategia es más común en los estudios de casos y controles y habría limitado esta revisión a una discusión de los pros y contras de una elección u otra, algo que se encuentra muy comentado en libros de texto. Además, la elección de dos grupos de referencia no siempre es recomendable³².

Según lo afirmado hasta ahora puede parecer que la discordancia entre los estudios de casos y controles hospitalarios y los comunitarios es la norma, cuando no es así, y hay muchísimos ejemplos en la literatura científica en los que no se han encontrado diferencias sustanciales entre ambos diseños, ni con los estudios de cohortes. Esto volvería a traer la pregunta de la frecuencia de discordancias entre los estudios hospitalarios y los comunitarios. Según lo afirmado al principio de este comentario creemos que es muy difícil responder a esta pregunta, por la imposibilidad de enumerar la base de estudios que han tratado la misma pregunta (el denominador) y por no reflejar los autores estas diferencias en el resumen de los repertorios informatizados, lo que facilitaría al menos el conocimiento del numerador. Para la valoración de la frecuencia se requeriría seleccionar una muestra aleatoria de investigaciones y establecer en ellas mediante un protocolo estructurado la posibilidad de que el error se presente, una tarea ingente y de dudosa rentabilidad en un ámbito general, aunque muy pertinente ante una pregunta de investigación específica. Por ello, hay que utilizar la estrategia de revisión narrativa, basada en la ilustración de una serie de ejemplos, más o menos característicos de preguntas de investigación concretas.

En la tabla 2 se resumen las razones para la falta de consistencia, así como las soluciones para remediar los posibles errores. Los ejemplos anteriores ponen de manifiesto que los estudios hospitalarios son más proclives al sesgo de selección que los estudios comunitarios, simplemente por trabajar con una población seleccionada y no con los casos que se generan en la colectividad. Los sesgos de selección que con mayor frecuencia se presentan en los estudios hospitalarios, cuando abordan problemas que se originan en la colectividad, son:

– Sesgo de referencia: la población que atiende el centro no representa lo que sucede en la colectividad. Este error tiene trascendencia cuando se intenta esta-

Tabla 2. Problemas más frecuentes y recomendaciones a seguir en la investigación hospitalaria

<i>Problema:</i> sesgo de referencia/remisión, especialmente agudo en los centros terciarios
<i>Solución:</i> analizar la remisión y la estadificación del proceso
<i>Problema:</i> uso de la historia clínica
<i>Solución:</i> reflexionar sobre cómo se ha completado. Tratar los «no consta» de manera separada. Lo mejor es que la recogida de datos no sea retrospectiva
<i>Problema:</i> los sesgos más frecuentes en los estudios de casos y controles hospitalarios (de base secundaria) son: los sesgos de Berkson y de inclusión
<i>Solución:</i> identificación correcta de la base del estudio (secundaria) y utilización de criterios de selección adecuados
<i>Problemas:</i> sesgo de detección en las enfermedades de largo período de latencia
<i>Solución:</i> analizar cómo se realiza el diagnóstico del efecto y comprobar que no se practica con más frecuencia en los expuestos
<i>Problema:</i> sesgo protopático
<i>Solución:</i> pensar siempre en el caso de enfermedades con fases prolongadas de latencia y seleccionar, si es posible, casos sin clínica
<i>Problema:</i> confusión por indicación
<i>Solución:</i> analizar las razones que motivan la aplicación de una intervención y tratarlas en el análisis. Realizar análisis en función del tiempo transcurrido entre la intervención y el efecto
<i>Problema:</i> falta de validez externa
<i>Solución:</i> no hay. Si la población estudiada no representa a la que se pretende aplicar los resultados, hay que hacer otro estudio

blecer la frecuencia de la enfermedad. Este tipo de error tiene también como consecuencia una frecuencia distorsionada de una exposición (en general, aumentándola). Si la exposición está aumentada, la consecuencia en la detección de un factor de riesgo, bajo el modelo de causalidad de Rothman, es que aumentan el valor del RR. Esto puede ayudar a explicar el porqué en muchos metaanálisis se aprecia que el valor de RR que dan los estudios hospitalarios es superior al de los estudios de la colectividad.

– Los sesgos típicos de los estudios de casos y controles hospitalarios: sesgo de inclusión, sesgo de exclusión y sesgo de Berkson. El sesgo de Berkson se conoce desde hace tiempo y es relativamente fácil evitarlo⁴⁻⁶. El sesgo de inclusión también es muy fácil de evitar, no obstante en uno de los ejemplos anteriores es la única explicación que se puede aducir para justificar la diferencia existente entre los diseños asentados en el hospital y los realizados con controles comunitarios. La solución para la lucha contra los sesgos de selección en los estudios de casos y controles pasan por una correcta identificación de la base de población (población secundaria) a la que intenta representar el estudio.

En teoría los sesgos de detección no son más frecuentes en los estudios de casos y controles que en otro tipo de diseños. No obstante, en dos de las situaciones que se han comentado es una explicación ra-

zonable para las discordancias observadas entre los estudios de población y los hospitalarios. Los pacientes que sufren vasectomía son mejor controlados desde el punto de vista médico, lo que puede facilitar que se diagnostique antes el cáncer de próstata. Lo mismo puede suceder con los cánceres de colon proximal, que son los más difíciles de diagnosticar por cribado, y ello puede facilitarse en los pacientes que se han operado de colecistectomía. Conviene, por lo tanto, estar en guardia ante procesos de largo período de latencia, en los que es posible un adelanto diagnóstico, y éste es más probable que ocurra cuando los casos se reclutan en el hospital. Hay que reflexionar sobre cómo se ha producido el diagnóstico del efecto en la población y si éste se ve influido por la exposición.

El sesgo protopático no es especialmente frecuente en los estudios anidados en el hospital, aunque en la situación propuesta (alcohol y litiasis biliar) un diseño basado en casos diagnosticados en el hospital de manera pasiva puede producirlo. La estrategia en estos casos es primero pensar en él, y luego, en caso de sospecha, hacer análisis con fracciones de los casos, en las que se desechan todos los enfermos que no hayan sido diagnosticados de manera activa a través de un cribado, o de manera accidental por problemas que no tienen la menor relación con la exposición. La confusión por indicación se produce sobre todo en el análisis de medicaciones o intervenciones, ambulatorias u hospitalarias. En estos casos hay que analizar las razones que motivan la intervención y realizar el análisis de acuerdo con ellas. El análisis en función de la latencia (tiempo transcurrido desde la exposición) también puede ayudar, si se comprueba que la fuerza de la asociación desciende conforme el tiempo de latencia aumenta.

Otro de los problemas que se ha suscitado es el que se refiere a la validez externa de la muestra (visto en la vacuna frente al neumococo). Las poblaciones hos-

pitalarias pueden no ser las adecuadas para generalizar medidas de intervención para la comunidad, ya que la muestra de sujetos de la colectividad que se utilizan tienen, en general, un riesgo más elevado de padecer cualquier proceso. Los procesos de aleatorización y enmascaramiento no sirven para controlar este error. Este problema no tiene solución: uno de los principios básicos de investigación es que la población de estudio debe representar a la que se pretende aplicar los resultados y hay que intentar cumplirlo.

Hay que recordar los problemas que plantea la historia clínica como herramienta única de obtención de información. Es recomendable en estos casos tratar los «no consta» como tales y no asumir que son «no expuesto». Si la frecuencia de los «no consta» es excesiva, nunca se estará seguro de que la asociación que se encuentre sea real. Los problemas de falta de uniformidad introducen un error de mala clasificación en principio no diferencial, que si la exposición tiene más de dos niveles, conviene recordar que puede sesgar el valor de la asociación en cualquier sentido³³. Lo mejor siempre es una recogida prospectiva de información en la que se minimicen las pérdidas y se mantenga el principio de uniformidad.

A modo de conclusión, para realizar una investigación hospitalaria que responda correctamente ante un problema que surge en la comunidad, aparte de una recogida de datos apropiada y uniforme, es conveniente plantear el marco de población en la que se origina, el proceso por el que un individuo llega a una institución, la forma en que se realiza el diagnóstico (influencias de la exposición u otros procesos relacionados con la exposición) y si las intervenciones se aplican como consecuencia de otros procesos relacionados con el efecto o por el propio efecto en sí. Con estos elementos los resultados observados en los estudios hospitalarios deberían ser más coincidentes con los procedentes de los estudios comunitarios.

Bibliografía

1. Hill AB. The environment and disease: association or causation? *Proc Royal Soc Med* 1965;58:295-306.
2. Berkson J. Limitations of the application of fourfold table analysis to hospital data. *Biomet Bull* 1946;2:47-53.
3. Feinstein AR, Walter SD, Horwitz RI. An analysis of Berkson's bias in case-control studies. *J Chron Dis* 1986;39:495-504.
4. Flanders WD, Boyle CA, Boring JR. Bias associated with differential hospitalization rates in incident case-control studies. *J Clin Epidemiol* 1989;42:395-401.
5. Robert RS, Spitzer WO, Delmore T, Sackett DL. An empirical demonstration of Berkson's bias. *J Chron Dis* 1978;31:119-28.
6. Walter SD. Berkson's bias and its control in epidemiologic studies. *J Chron Dis* 1980;33:721-5.
7. Delgado Rodríguez M. Sesgos en el estudio de factores pronósticos. *Med Clin (Barc)* 1999;112(Supl 1):51-8.
8. Horwitz RI, Stewart KR. Effect of clinical features on the association of estrogens and breast cancer. *Am J Med* 1984; 76:192-8.
9. Romano PS, Roos LL, Luft HS, Jollis JG, Doliszny K, and the Ischemic Heart Disease Patient Outcomes Research Team. A comparison of administrative versus clinical data: coronary artery bypass surgery as an example. *J Clin Epidemiol* 1994; 47:249-60.
10. Delgado-Rodríguez M, Gómez-Olmedo M, Bueno-Cavanillas A, García-Martín M, Gálvez-Vargas R. Recall bias in a case-control study of low birth weight. *J Clin Epidemiol* 1995;48: 1133-40.
11. Sackett DL. Bias in analytic research. *J Chron Dis* 1979;32: 51-63.
12. Hulley SB, Cummings SR, editors. *Designing clinical research: an epidemiologic approach*. Baltimore: Williams & Wilkins, 1987.

13. Schoenberg BS, Anderson DW, Haerer AF. Prevalence of Parkinson's disease in the biracial population of Copiah County, Mississippi. *Neurology* 1985;35:841-5.
14. Feinstein AR. *Clinical epidemiology. The architecture of clinical research*. Philadelphia: Saunders, 1985.
15. Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Stat Med* 1997;16:981-91.
16. Rothman K. Causes. *Am J Epidemiol* 1976;104:587-92.
17. Morgenstern H, Glazer W, Niedwiecki D, Nourjah P. The impact of neuroleptic medication on tardive dyskinesia: a meta-analysis of published studies. *Am J Public Health* 1987;77:717-24.
18. Bernal-Delgado E, Latour-Pérez J, Pradas-Arnal F, Gómez-López LI. The association between vasectomy and prostate cancer: a systematic review of the literature. *Fertil Steril* 1998;70:191-200.
19. Giovannucci E, Colditz GA, Stampfer MJ. A meta-analysis of cholecystectomy and risk of colorectal cancer. *Gastroenterology* 1993;105:130-41.
20. Huang J-Q, Sridhar S, Chen Y, Hunt RH. Meta-analysis of the relationship between *Helicobacter pylori* seropositivity and gastric cancer. *Gastroenterology* 1998;114:1169-79.
21. Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. *Am J Epidemiol* 1999;149:981-3.
22. Thijs C, Knipschild P, Leffers P. Does alcohol protect against the formation of gallstones? A demonstration of protopathic bias. *J Clin Epidemiol* 1991;44:941-6.
23. Blais L, Ernst P, Suissa S. Confounding by indication and channeling over time: the risks of β_2 -agonists. *Am J Epidemiol* 1997;144:1161-9.
24. Johnson AG, Jick SS, Perera DR, Jick H. Histamine-2 receptor antagonists and gastric cancer. *Epidemiology* 1996;7:434-6.
25. Weinberg CR, Baird DD, Wilcox A. Sources of bias in studies of time to pregnancy. *Stat Med* 1994;13:671-81.
26. Sackett DI, Whelan G. Cancer risk in ulcerative colitis: scientific requirements for the study of prognosis. *Gastroenterology* 1980;78:1632-5.
27. Fine MJ, Smith MA, Carson CA, Meffe F, Sankey SS, Weissfeld LA, et al. Efficacy of pneumococcal vaccination in adults. *Arch Intern Med* 1994;154:2666-77.
28. Delgado-Rodríguez M, Sillero-Arenas M, Matín-Moreno JM, Gálvez Vargas R. Oral contraceptives and cancer of the cervix uteri. A meta-analysis. *Acta Obstet Gynecol Scand* 1992;71:368-76.
29. Steinberg KK, Smith SJ, Thacker SB, Stroup DF. Breast cancer risk and duration of estrogen use: the role of study design in meta-analysis. *Epidemiology* 1994;5:415-21.
30. Coughlin SS, Giustozzi A, Smith SJ, Lee NC. A meta-analysis of estrogen replacement therapy and risk of epithelial ovarian cancer. *J Clin Epidemiol* 2000;53:367-75.
31. Ben-Shlomo Y, Markowe H, Shipley M, Marmot MG. Stroke risk from alcohol consumption using different groups. *Stroke* 1992;23:1093-8.
32. Lasky T, Stolley PD. Selection of cases and controls. *Epidemiol Rev* 1994;16:6-17.
33. Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value. *Am J Epidemiol* 1990;132:746-8.
34. Delgado Rodríguez M. Unidad Didáctica 1. Investigación Científica. En: Doménech JM, editor. *Diseño de estudios sanitarios*. Barcelona: Signo, S.A., 2002; p. 1-126.