

Estudios de casos y controles en psiquiatría: causalidad, diseño y advertencias

L. C. Silva Ayçaguer

Instituto Superior de Ciencias Médicas de La Habana. Cuba

Case-control studies in psychiatry: causality, design and warnings

Resumen

El presente trabajo es de naturaleza esencialmente metodológica y se ha escrito con la expectativa de ayudar a que la investigación en psiquiatría se lleve adelante con creciente calidad, así como para contribuir a que los lectores de trabajos ajenos dispongan de pautas valorativas adecuadas. Se reflexiona brevemente en torno a las condiciones más importantes que necesariamente han de cumplirse para probar una hipótesis de causalidad, independientemente del diseño de investigación de que se trate. Sin embargo, el núcleo central del texto se destina a examinar e ilustrar el problema que tales condiciones plantean a quien diseñe y conduzca un estudio de casos y controles, que constituye la expresión dominante en la investigación observacional contemporánea. Además de reseñar los aspectos básicos en materia de diseño y análisis, áreas profusamente ilustradas con ejemplos de la literatura actual sobre casos y controles en psiquiatría, se ofrece un conjunto de advertencias sobre aspectos que suelen no contemplarse y cuyo olvido deteriora y ocasionalmente invalida las investigaciones desarrolladas con este importante método.

Summary

This present paper is mainly methodological and has been written with the aim of helping researchers in psychiatry to produce results with higher quality and help readers to have adequate assessment values of others. Brief reflection is made on the most important conditions that must be fulfilled to prove a causality hypothesis, regardless of the investigation design used. However, the main purpose of the text is to examine and illustrate how these conditions work under a case-control study environment. Besides outlining the basic aspects concerning design and analysis, areas extremely illustrated with examples of case and controls in psychiatry found in the literature, a number of suggestions to avoid pitfalls that can invalidate research efforts developed using case-control methodology is offered.

Key words: Case-control studies. Biases. Stepwise regression. Prediction. Logistic regression. Sample size.

Palabras clave: Estudios de casos y controles. Sesgos. Regresión paso a paso. Predicción. Regresión logística. Tamaño muestral.

1. INTRODUCCIÓN

Por lo general, la investigación clínica y epidemiológica se emprende, bien con el fin de describir una realidad o proceso (caracterizar en relativamente pocos trazos cómo es), bien para explicar por qué dicha realidad es como es, lo cual equivale en esencia a descubrir o corroborar los mecanismos causales que la gobiernan.

Varias condiciones han de cumplirse necesariamente para «tener derecho» a creer que cierta condición influ-

ye causalmente en otra. Para aludir a una fuente donde se enumeran suele citarse la célebre lista de condiciones de Hill¹, la cual ha sido complementada con otras demandas². Sin embargo, no todas estas exigencias tienen igual jerarquía. Por ejemplo, una premisa medular es que la hipótesis causal sea plausible. Se trata de un requisito anterior a cualquier acción metodológica e independiente de ella; constituye el primer tamiz ineludible. Esto quiere decir que, además de contar quizá con cierto aval empírico, debe tener un respaldo teórico racional, y en cualquier caso no puede estar en contradicción directa con hechos establecidos por la ciencia. Tal exigencia dista de ser un adorno lateral. La ausencia (o la vaga formulación) de una fundamentación relativamente persuasiva que confiera plausibilidad a la hipótesis no es un fenómeno infrecuente. Lamentablemente, los estudios concebidos a partir de anécdotas más o menos espectaculares, muchas veces relacionadas con energías indefinidas que nadie ha podido medir, no son pocos.

Correspondencia:

Luis Carlos Silva Ayçaguer
Instituto Superior de Ciencias Médicas de La Habana
Vicerrectoría de Investigaciones
Calle G y 25, 6.º piso, plaza
Ciudad de la Habana (Cuba)
Correo electrónico: lcsilva@infomed.sld.cu

Un ejemplo extremo pero, quizá por eso mismo, elocuente, lo ofrece en mi opinión la homeopatía³, modalidad terapéutica que ha despertado algún interés en el marco de la psiquiatría. En un artículo en que se exalta el papel de tal recurso terapéutico en el tratamiento de la fobia social, publicado en un *Journal* recientemente fundado para difundir estudios sobre las llamadas terapias alternativas, Davidson y Gaylord⁴ reconocían que «no hay ninguna explicación aceptable para la homeopatía dentro del marco de la teoría médica actual», y nos recuerdan una doctrina que se proclama desde su propio título en otro artículo de esa misma revista (*El mecanismo de la homeopatía: lo único que importa es que funciona*)⁵. Estimo que, salvo que sobren las finanzas o que otras circunstancias sociales lo aconsejen, carece de mayor sentido ocuparse de examinar recursos terapéuticos que, como este, no sólo carecen de toda fundamentación, sino que difícilmente la alcanzarán alguna vez, ya que sus cultores (acaso sabiendo que no la hallarían) se jactan de no necesitarla.

En el presente trabajo, sin embargo, pondré el énfasis en el examen de aquellas premisas para la causalidad que están directamente relacionadas con aspectos metodológicos, especialmente en aquellas que, a mi juicio, son verdaderamente cruciales e insoslayables; a saber: existencia de asociación, congruencia temporal, ausencia de sesgos responsables de la asociación observada, y que esta última no se pueda explicar por la acción de otros factores (los llamados factores confusores). Cualquiera que sea el diseño que se emplee, ha de lidiarse con las cuatro condiciones arriba enumeradas. El apartado 2 de este trabajo se destina a comentar someramente cada una de ellas.

En cualquier ámbito del conocimiento médico, y por ende en el campo de la psiquiatría, los recursos metodológicos en materia de diseño para corroborar conjeturas etiológicas o, más generalmente, hipótesis causales, recorren un amplio abanico de variantes y subvariantes. No obstante, consignados de manera concisa y esencial, dichos procedimientos son básicamente tres: los estudios de casos y controles, los longitudinales o de cohorte y los experimentales. Estos últimos, de los cuales el ensayo clínico controlado es su expresión más emblemática en el mundo biomédico, constituyen la forma óptima para encarar el problema de la causalidad. El carácter prospectivo inherente a todo experimento resuelve el escollo posible de la temporalidad, y si el ensayo clínico se conduce adecuadamente (en particular cuando se usan técnicas de enmascaramiento y asignación aleatoria) muchos de los posibles sesgos y el efecto de terceros factores son conflictos que quedan conjurados por el propio diseño.

Lamentablemente, imperativos éticos y prácticos de índole diversa suelen imposibilitar la experimentación. Los estudios observacionales de cohorte son el sucedáneo natural del ensayo clínico, ya que comparten con él su naturaleza prospectiva: los hechos se registran en orden temporalmente ascendente, la observación comienza tanto para quienes tienen una presunta condición de riesgo como para quienes no la tienen, cuando aún no se han expresado los desenlaces que se estudian y, en

esencia, se trata de comparar las tasas de incidencia entre dichos grupos, de modo que, salvo situaciones excepcionales (v. artículo de Silva y Benavides⁶), el registro adecuado de la temporalidad queda garantizado de antemano. Pero este enfoque alarga los lapsos para llegar a resultados analizables (a veces durante años) y suele ser prohibitivamente caro.

Tales circunstancias hacen de los estudios de casos y controles el recurso adecuado cuando los recursos son escasos y el tiempo disponible reducido; y acaso sea esa la razón que los ha convertido en la forma dominante de la investigación etiológica contemporánea. Este enfoque, sin embargo, no es una panacea: sólo cuando se consiguen superar las diversas acechanzas metodológicas que le son propias puede producir resultados de alta calidad, similar incluso a la de los que se generan por conducto de los ensayos clínicos (v. la serie de trabajos publicados recientemente en *Lancet* debidos a Schulz y Grimes, especialmente el que se dedica a este tipo de estudios⁷). En el apartado 3 se exponen los rasgos básicos que definen dicho procedimiento. Procuro reflexionar sobre los aspectos más relevantes en materia de diseño, así como bosquejar las líneas de análisis fundamentales que le corresponden. Sin embargo, puesto que no falta bibliografía especializada sobre el tema, incluyendo libros^{8,9} y números enteros de revistas (p. ej., un número monográfico de *Epidemiologic Reviews**), el tratamiento del tema en esta sección apunta al manejo de los aspectos más generales y se desarrolla insistiendo más en lo conceptual que en lo operativo.

El apartado 4, finalmente, se concentra en compartir algunas advertencias que, si bien conciernen a errores relativamente fáciles de evitar, es menester tener en cuenta para no padecer consecuencias que pueden ser catastróficas, como quedará debidamente ilustrado. Se trata, por lo demás, de precauciones insuficientemente explicadas de modo razonado y comprensible en la literatura no especializada.

Nuestra expectativa no sólo es dar pautas o ideas que propendan a que la investigación en psiquiatría incremente su excelencia, sino también contribuir al desarrollo de la capacidad crítica de los lectores de artículos científicos, en particular de aquellos numerosos trabajos en que se da cuenta de resultados obtenidos con la metodología de casos y controles.

2. PRINCIPALES CONDICIONES PARA LA CAUSALIDAD

2a. Asociación

La existencia de asociación es una premisa evidente de la causalidad. Pero en la práctica, y pese a reiteradas advertencias, se confunde a diario con la causalidad propiamente dicha. No es infrecuente tropezar con proyectos,

* Se trata del número 1 del volumen 16 publicado en 1994.

tesinas e incluso artículos publicados en prestigiosas revistas entre cuyos objetivos se halla el de identificar la asociación existente entre dos variables. Las fórmulas empleadas para tal declaración varían: en algunos casos se anuncia el propósito de «evaluar si hay correlación», otras comunican que se desea «establecer la asociación existente», y algunas plantean directamente que se quiere «correlacionar» una variable (típicamente, un factor presuntamente causal) con otra (por lo general un «desenlace» de cierto proceso). Lamentablemente, así planteado, esto carece de sentido. El acto de computar un coeficiente de asociación no entraña más que la aplicación mecánica de un método y no puede ser, por ende, un genuino objetivo de investigación. Está muy extendido el temor a admitir franca y claramente que se quiere «probar que X influye en Y» o «evaluar el grado en que X es causa de Y» y se opta por el subterfugio de comunicar que sólo se quiere «cuantificar la asociación entre X y Y» como si la asociación pudiera tener un interés intrínseco. Se suple simplemente la finalidad con uno de los medios para conseguirla.

2b. Congruencia temporal

Todos sabemos que un vínculo causal presupone una relación temporal en el sentido de que la presunta causa de cierto efecto, necesariamente, ha de precederlo en el tiempo. Desde muy niños advertimos que la iluminación de un salón se asocia con que se haya accionado un interruptor, pero —debido al orden en que inexorablemente se producen dichos acontecimientos— también aprendemos rápidamente que es esta última acción la que produce la luz y no viceversa. Ningún análisis de causalidad tiene un sentido claro cuando en el diseño del estudio no se ha adoptado la precaución de haber registrado los acontecimientos de suerte que se pueda luego contemplar una regla tan básica como la enunciada. El peligro de no respetarla es muy notable en los estudios de casos y controles. En el subapartado 3b se podrá apreciar que ocurre con frecuencia nada desdeñable.

2c. Efecto de factores de confusión

Una explicación posible para el hecho de haber observado una asociación entre dos variables es que ella sea mero reflejo de la acción de un tercer factor sobre ambas; en tal caso nuestra observación podría ser atribuida a una relación de naturaleza estructural a pesar de tratarse de algo simplemente fenoménico. Para no incurrir en este error interpretativo, en estudios no experimentales, tales como los de casos y controles, es necesario apelar a técnicas especiales como las que se bosquejan en el subapartado 3c.

2d. Sesgos

Se ha dicho que realizar una buena investigación se reduce a evitar los sesgos que acechan a todo diseño. Tal

es el caso, por poner un ejemplo, del llamado sesgo del interrogador, consistente en que quien interroga o mide a los participantes, por la razón que sea, no pone igual celo en la captación de información para todos los elementos (es menos cuidadoso en el registro de datos de los sanos que de los enfermos, o más inquisitivo si se trata de jóvenes que de ancianos). El inventario de sesgos posibles es cuantioso (de anticipación en el diagnóstico, del interrogador, de memoria, de selección, etc.) y la literatura que los aborda, igualmente copiosa. Para acceder a una descripción detallada y profusamente ilustrada recomiendo un reciente y acucioso libro de Moisés Szklo y Javier Nieto¹⁰. En el subapartado 3d se tratan en detalle los dos sesgos más importantes que suelen aquejar a los estudios de casos y controles.

3. DISEÑO DE CASOS Y CONTROLES

Hecho este repaso general, y expuestas en este contexto las cuatro premisas básicas para convalidar una hipótesis de causalidad, consideraremos en esta sección los problemas específicos que plantea el manejo de cada una de ellas a quien opere con un estudio de casos y controles.

El rasgo fundamental de esta metodología consiste en que procura resolver el problema de la causalidad mediante un proceso de reconstrucción de los acontecimientos al final del cual se han de comparar dos grupos: el de los individuos que han tenido un desenlace específico y el de quienes han transitado por un proceso similar, pero han desembocado en el desenlace opuesto. La frecuencia con que se registraron los factores que se sospechan causalmente relacionados con dicho desenlace constituye el objeto de la comparación entre los dos grupos.

Dicho de otro modo, en estos estudios un grupo de individuos (los casos) portadores de una condición dada (usualmente una enfermedad, pero ocasionalmente otro tipo de suceso adverso como, por ejemplo, un intento suicida) son elegidos para ser comparados con un grupo de sujetos (los controles) que no exhiben dicha condición. Para unos y otros se registran datos pasados (con frecuencia llamados «factores de exposición») que se suponen relevantes a los efectos de desembocar en la condición morbosa que se estudia, tales como pudieran ser el peso al nacer, el consumo de drogas o determinada práctica sexual.

Estos estudios deben su notable popularidad básicamente a la agilidad y economía con que pueden desarrollarse. Constituyen la alternativa natural a los estudios de cohorte, que son más caros, de mayor duración y más complejos desde el punto de vista logístico. A lo largo de la primera mitad del siglo xx se verificó la evolución fundamental del método, sobre todo en aras del deseo de elucidar la etiología de las enfermedades crónicas en una sociedad altamente industrializada, cuyo patrón de morbilidad y mortalidad ya no respondía al marco de las infecciones. En ese contexto surgió el grueso de los métodos estadísticos con que se han venido manejando los

datos procedentes de este tipo de estudios. Una de las más importantes contribuciones la debemos a Cornfield¹¹, quien sugirió emplear la razón de *odds ratio* (OR) para estimar el riesgo relativo en una investigación de casos y controles. En el Reino Unido el gran impulsor de este procedimiento fue Bradford Hill, quien cobró fama a partir del célebre estudio realizado en 1950 con su alumno Richard Doll¹², donde se valoraba el efecto del hábito de fumar sobre el cáncer de pulmón.

3a. Asociación

Coherentemente con el hecho de que carece de sentido proponerse la identificación de correlaciones en sí mismas, a nadie se le ocurriría investigar la asociación entre el hecho de que un adolescente intente suicidarse y, pongamos por caso, el color de la pared del cuarto en que pernocta. ¿Por qué? Simplemente porque aunque la asociación entre dos condiciones sea una condición *sine qua non* para concebir que estén causalmente vinculadas, siempre que se procura constatar su existencia es debido a que se tiene una sospecha (como mínimo subconsciente) de que tal corroboración pudiera brindar una prueba (o al menos un indicio) de una relación causal¹³.

Por ejemplo, cuando Zornberg y Jick estudian la relación entre el consumo de neurolépticos y la aparición de tromboembolismo venoso idiopático¹⁴ no se limitan a computar una medida de asociación (digamos, una razón de OR) y a realizar una prueba de hipótesis para confirmar que dicha asociación no es exclusivamente atribuible al azar. De hecho, lo que se preguntan es si ese consumo incrementa el riesgo de la dolencia, de modo que adoptan todo un conjunto de precauciones metodológicas que permitan trascender la mera constatación fenomenológica de que las variables exhiben asociación.

Aunque el propósito de este trabajo es básicamente conceptual, resulta imposible no introducir, siquiera brevemente, algunos elementos matemáticos básicos. En el entorno de los estudios de casos y controles el procedimiento por antonomasia para medir asociación es, sin duda alguna, la OR.

Para ilustrarlo supongamos que se estudió un conjunto de 181 adolescentes que hicieron un intento suicida y que por cada uno de ellos se estudiaron tres adolescentes de igual edad y sexo elegidos de la misma comunidad, los cuales nunca atentaron contra su vida. Imaginemos que el porcentaje de los que tuvieron una pérdida significativa de peso durante el semestre anterior a la fecha del acto suicida ascendió al 70% entre los primeros y a sólo un 56% entre los segundos. Hemos observado una clara asociación entre intento suicida y pérdida de peso. Los datos mencionados pueden resumirse como puede apreciarse en la tabla 1.

La OR se computa o estima mediante la llamada «razón de productos cruzados»:

$$OR = \frac{ad}{bc} = \frac{(128)(241)}{(53)(302)} = 1,9$$

TABLA 1. Distribución de casos y controles según hubiesen tenido o no pérdida de peso antes del intento suicida

Intento suicida	Pérdida de peso		Total
	Sí	No	
Sí (casos)	a = 128	b = 53	181
No (controles)	c = 302	d = 241	543
Total	430	294	724

En este ejemplo, en principio, puede considerarse que es casi dos veces más probable que un chico que ha perdido peso realice un intento suicida a que lo haga uno que conserva el peso normal*. Nótese en la fórmula que el numerador de la OR será grande cuando a y d lo sean y el denominador pequeño cuando sean pequeños c y b. Consecuentemente, la OR tiende a ser mayor en la medida que los datos tiendan a ubicarse en la primera diagonal de la tabla, indicando así mayor asociación entre el presunto factor de riesgo (pérdida de peso) y el desenlace (intento suicida). Para valorar si este número es «significativamente» mayor que la unidad, hasta hace relativamente poco era usual que se llevara adelante una prueba de significación que da lugar a un valor de chi cuadrado y un valor de *p* asociado. Según las reglas entronizadas durante decenios, si el valor de *p* es inferior a 0,05, entonces se considera que hay «significación»; es fácil corroborar por medio de cualquiera de los programas informáticos disponibles al efecto que si optáramos por este enfoque obtendríamos chi cuadrado = 12,8 y *p* = 0,0003. Este procedimiento en general, pero especialmente en el análisis de las OR, ha ido cayendo en franco desuso (v. el subapartado 4b). Desde la aparición de un contundente artículo de Gardner y Altman¹⁵ en el *British Medical Journal*, el uso de la prueba de hipótesis se ha venido supliendo con la comunicación de los límites de confianza para la OR. Al igual que antes, es muy fácil hoy hallar recursos informáticos que permiten hacer tal cómputo. En el ejemplo tendríamos para esta OR que el intervalo (al 95 % de confiabilidad) es 1,3-2,8, lo cual nos conduce a afirmar que podemos estar altamente confiados en que la OR es como mínimo 1,3, aunque podría ascender a 2,8. Es decir, la probabilidad de un intento suicida es, cuando menos, un 30% mayor para un adolescente que haya tenido una pérdida de peso sensible en el último semestre que para uno que no haya estado en ese caso.

3b. Congruencia temporal

No tener en cuenta la premisa de precedencia temporal es una trampa abierta, especialmente insidiosa en los

*Típicamente, la OR puede considerarse igual al riesgo relativo, un parámetro que sólo se puede estimar en los estudios de cohorte (salvo que se cuente con información adicional).

estudios transversales y retrospectivos, en los que se debe indagar sobre hechos ocurridos con anterioridad al momento del estudio. La clave del problema radica en que es imposible en estos casos establecer mediante observación cuál fue el orden en que ocurrieron los hechos que se registran. Los estudios de casos y controles donde sólo puede intentarse la reconstrucción de los acontecimientos son altamente vulnerables a la aparición de este problema.

En los estudios retrospectivos, además de sujetos sanos, se incluyen individuos que en el momento de la encuesta padecen una enfermedad o tienen una condición dada. Por ejemplo, los casos podrían ser mujeres que padecen anorexia nerviosa y los controles mujeres sanas en ese sentido. Supongamos que para cada una de ellas (casos y controles) se obtienen datos tales como antecedentes familiares de trastornos mentales, si consume anfetaminas y si practica ejercicio. El antecedente de una madre o un padre con desórdenes psiquiátricos es quizás un hecho muy probablemente anterior a la situación que presente esta mujer ahora (la dolencia paterna pudo haber sido diagnosticada incluso antes del nacimiento del hijo). Pero para evaluar el posible efecto causal, por ejemplo, del consumo de anfetaminas o del sedentarismo en el desarrollo de la enfermedad lo que realmente interesa registrar no es si el sujeto consume dichos fármacos o practica ejercicio en el momento de la encuesta o poco antes, sino si lo hacía o no durante un período previo a la aparición de la dolencia suficientemente largo como para haber tenido la oportunidad de producirla.

Tratando de evaluar si ciertos factores de riesgo son específicos para la anorexia nerviosa o válidos más generalmente para otros desórdenes psiquiátricos generales o de la alimentación, recientemente se estudiaron 67 mujeres con diagnóstico de anorexia nerviosa y 102 controles con otros desórdenes psiquiátricos, así como otros 102 con bulimia nerviosa¹⁶. Los investigadores sondean un amplio espectro de factores de riesgo, tales como obesidad de la madre, trastornos mentales de los padres, menarquia temprana y personalidad perfeccionista. Obsérvese que todos estos ejemplos aluden a rasgos que parecen razonablemente suponerse anteriores al eventual desarrollo de la enfermedad. Consecuentemente, en este estudio las posibles asociaciones obtenidas entre estos factores y la anorexia supera esta exigencia, propia de una interpretación causal. Hay que reparar en que en los estudios socioepidemiológicos ocurre muy frecuentemente que los fenómenos se «retroalimentan» mutuamente como causa y efecto. Por ejemplo, puede ser que la depresión sea un factor de riesgo para el desarrollo de la anorexia, pero ello no es óbice para que se produzca también la relación inversa: que como consecuencia de su anorexia el paciente desarrolle un síndrome depresivo o se agudice su presencia*. Quiere esto decir que si

los investigadores hubieran incluido la depresión como factor de riesgo tendrían que haberse asegurado de que la evaluación de tal condición se hubiese hecho antes del comienzo de la anorexia.

Cuando no se tiene en cuenta este «detalle» se pierde la lógica temporal del estudio y se cancela toda interpretación potencial de sus resultados, hecho que puede ser desastroso, aunque muchos investigadores no lo tomen en cuenta o lo consideren como un mal menor que se «resuelve» consignándolo como una limitación del estudio.

En un artículo muy reciente¹⁷ se indaga sobre el uso de drogas, alcohol y tabaco por parte de pacientes con diagnóstico de esquizofrenia en Escocia. Para ello de determina por interrogatorio el consumo de esas sustancias por parte de 316 esquizofrénicos y de 250 sujetos tomados de la población general. Como conclusión, los autores dan cuenta de que «más pacientes comunicaron el uso de drogas el año previo (7 frente a 2%) y mayor consumo de alcohol (17 frente a 10%)». Puesto que, obviamente, la enfermedad no comenzó el día del interrogatorio, sino probablemente varios años antes, estos porcentajes serían más bien tasas de incidencia de drogadicción y alcoholismo entre enfermos y entre sanos. Lamentablemente, ni siquiera de eso puede estarse seguro, pues para algunos de los consumidores el hábito pudiera haberse establecido antes de aparecer la enfermedad, e incluso haberla favorecido. Haciendo algo más confusas las cosas, los autores agregan: «Los fumadores actuales fueron un 65% de los enfermos frente a un 40% en la población», y luego concluyen que «el problema de uso de drogas y alcohol es mayor entre esquizofrénicos que en la población general». Sorprende hallar estos resultados en un artículo cuyo título (*Uso de drogas, alcohol y tabaco por esquizofrénicos: estudio de casos y controles*) alude directamente a un estudio retrospectivo, aunque a la postre comunique lo contrario (aparición de factores entre los enfermos).

Examinemos con cierto detalle otro ejemplo. En un artículo titulado *Depresión y demencia: estudio caso-control*^{**} los autores afirman que se proponen conocer los factores de riesgo de la depresión en pacientes con deterioro cognitivo¹⁸. Sus conclusiones, sin embargo, son las siguientes: «el deterioro cognitivo, los antecedentes psiquiátricos y los factores de riesgo o enfermedad cerebrovascular influyen en la presencia de depresión en muestras geriátricas. Sexo, edad, estado civil, subtipo, gravedad y duración de la demencia no se asocian al diagnóstico de depresión». Lamentablemente observo varios de los problemas que a mi juicio han de evitarse. En principio identifico los siguientes:

- Se dice que es un estudio de pacientes con deterioro cognitivo y luego se informa que el deterioro cognitivo es un factor de riesgo.

*Tal bidireccionalidad puede estar presente en los binomios demencia-alcoholismo, drogadicción-disfunción sexual, fracaso escolar-fobia social, etc.

**Incidentalmente llamo la atención al lector sobre el hecho de que a lo largo del texto he eludido la expresión «estudio caso-control» que considero una traducción mimética del inglés *case-control study*.

- Para algunas de las variables se «concluye» que ellas no están asociadas a la depresión (en lugar de valorar si son o no factores de riesgo como se anunciaba en los objetivos declarados).
- No queda nada claro cuáles son los «factores de riesgo» que influyen en la presencia de depresión.
- La conclusión alude a «muestras», lo cual carece de sentido, pues las preguntas nunca son realmente sobre las muestras, sino que conciernen a las poblaciones que ellas representan¹³.
- El manejo del tiempo es sumamente confuso y en cualquier caso improductivo: los autores estudian las relaciones causales entre demencia y depresión, pero, según comunican, sólo se conocía cuándo se inició la enfermedad para el 70% de los dementes, mientras que en el resto la duración recorrió un espectro que iba desde menos de 1 año (25%) hasta más de 6 años (10%), y otro tanto ocurrió con la duración de la depresión (inferior a 1 año en el 41%, superior a 5 años en un 22% y desconocida en un 7%).

El peligro de no reparar en esta crucial exigencia metodológica es mayor cuando se estudian enfermedades crónicas: el conocimiento de lo que ocurrió antes de su comienzo puede ser muy difícil (o imposible), simplemente debido a la dificultad (o imposibilidad) para identificar el momento en que comenzó el trastorno. En el examen de problemas agudos o puntuales, sin embargo, como ocurre con el suicidio, todo se reduce a adoptar las debidas precauciones; por ejemplo, Kresnow¹⁹ indaga acerca de exposiciones tales como alcoholismo, desesperanza, incitaciones indirectas al suicidio (por conducto de los medios de comunicación, amigos o parientes) y la depresión; los datos se registran, dependiendo de lo que se trata, para el año, el mes o el día anterior al intento suicida cuando se trata de los casos, y para los mismos lapsos, pero previos al momento del interrogatorio para los controles.

Para concluir con estas consideraciones sobre la temporalidad cabe señalar la notable confusión que suele producirse con el concepto de casos y controles. Un ejemplo muy reciente, tomado de una prestigiosa revista, lo pone de manifiesto²⁰. Se estudian tasas de disfunción sexual (variable sobre la que se indaga mediante un cuestionario autoadministrado) en 135 personas con esquizofrenia y en 114 sujetos elegidos de la población general. Se obtienen resultados tales como que entre los enfermos para el 82% de los hombres y para el 96% de las mujeres se halló al menos una disfunción sexual, los enfermos presentaron en mayor porcentaje problemas de erección y las mujeres menos disfrute que las mujeres sanas. Concluyen que «las personas con esquizofrenia comunican tasas mucho más altas de disfunción sexual que los que proceden de la población general». Pese a que en el mismísimo título se anuncia que han realizado un «estudio de casos y controles», parece que lo que se registra, tanto en individuos sanos como en sujetos que ya tenían la dolencia, es la presencia o ausencia actual de disfunciones. Tal confusión probablemente sea debida parcialmente a la

anfibología del término controles que se ha empleado genéricamente para aludir a los sujetos con los que se compara; sin embargo, es obvio que no es legítimo caracterizar el estudio con una denominación que está universalmente reservada para estudios retrospectivos.

3c. Factores de confusión

Supongamos que en un estudio de casos y controles se observa cierto valor de la OR; en principio ello nos permite pronunciarnos acerca de si hay o no una asociación entre las variables dicotómicas involucradas y, más generalmente, acerca de su intensidad. Sin embargo, la magnitud observada pudiera explicarse (total o parcialmente) por el efecto de un tercer factor que concommita con ambos; en tal caso se dice que este último es un factor de confusión.

Para que una variable pueda considerarse un factor de confusión de la asociación entre una exposición (o factor de riesgo) y una respuesta (o desenlace) tiene que cumplir formalmente lo siguiente: estar a la vez asociada con el desenlace y con la exposición, pero sin ser una consecuencia de esta última²¹; es decir, el posible factor de confusión para serlo no puede ser un eslabón causal intermedio entre exposición y desenlace (fig. 1).

El efecto de un factor de confusión puede ser tanto el de incrementar como el de atemperar la apreciación que tengamos de la asociación intrínseca o estructural que realmente vincula a las variables objeto de estudio. Ésta es la razón que aconseja manejar los datos de manera que, en la medida de lo posible, podamos conocer la magnitud del grado de asociación intrínseca que tengan dichas variables. Cuando se lleva adelante el proceso que lo permite se suele decir que «se ha controlado» el factor de confusión o que se ha obtenido una estimación de la OR «ajustada» por dicho factor.

En el ejemplo resumido en la tabla 1, la asociación observada pudiera explicarse por el hecho de que la presencia de síndrome depresivo en el semestre anterior es mucho más acusada entre quienes a la postre intentaron el suicidio que entre los que no lo hicieron, unido al de que la tasa de depresión es también mayor entre los que perdieron peso que entre quienes mantuvieron su peso normal o aumentaron de peso. En tal caso se dice que la asociación entre pérdida de peso e intento suicida está confundida por la depresión.

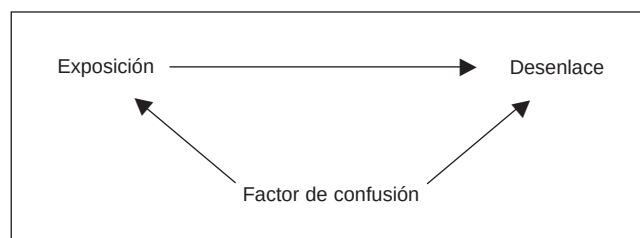


Figura 1. Relaciones entre causa, efecto y un factor de confusión.

Esto quiere decir que la posible condición de riesgo que se atribuía al hecho de que el adolescente hubiese perdido peso resulta dudosa, ya que se sospecha que la aparición de síntomas de depresión pudiera ser responsable de la asociación observada. Es decir, no se descarta que si se examina dicha asociación una vez «controlado» el efecto de la depresión ésta pudiera desaparecer, revelándose así que la pérdida de peso no sería en sí misma una condición influyente en la propensión al suicidio, aunque la asociación observada sugiriera en principio un vínculo causal.

Supongamos que al examinar los datos de la tabla 1, «cruzando» por una parte la condición de caso y control con el síndrome depresivo y por otra la de esta última condición con la pérdida de peso, obtenemos los resultados que se muestran en la tabla 2.

En esta relación la presencia de un síndrome depresivo cumple con los requisitos para ser considerada una variable que confunde la relación que vincula a la pérdida de peso con el intento suicida. En primer lugar es bien conocido que el riesgo de suicidio es mayor entre los que están deprimidos (aparte de que los datos del primer cruzamiento lo refrenda, pues

$$OR = \frac{(144)(332)}{(211)(37)} = 6,12$$

es un número mucho mayor que la unidad). Por otra parte, la depresión está fuertemente asociada con la pérdida de peso: sólo el 5 % de los no deprimidos bajó de peso, mientras que tal condición se dio para el 88 % de los que presentaban el síndrome. Y, por último, no tiene mayor sentido suponer que el supuesto efecto de la depresión se realice a través de la pérdida de peso (fig. 2).

En un trabajo archiconocido, titulado *Aspectos estadísticos del análisis de datos procedentes de estudios retrospectivos de la enfermedad*, Mantel y Haenszel²² introdujeron el recurso que imperó durante muchos años para encarar la tarea de «controlar» los factores confusores: la estratificación. La idea clave consiste, en efecto, en que cuando estamos antes un posible confusor éste debe y puede ocasionalmente «controlarse». Esto quiere

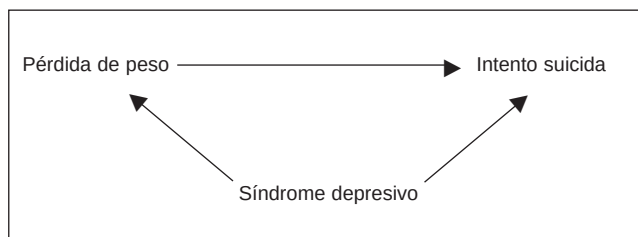


Figura 2. Relación entre la pérdida de peso, intento suicida y síndrome depresivo.

decir que es posible hacer una maniobra en el análisis que permite aquilatar el «efecto puro» de la presunta causa (tal y como si todos los sujetos de la muestra fueran iguales en lo que se refiere al factor de confusión). El resultado de dicha maniobra es la OR de Mantel y Haenszel (OR_{MH}), que no es más que un promedio de las OR observadas en los estratos*.

Para apreciar la idea, consideremos tres estratos considerando los sujetos no deprimidos: los que estuvieron moderadamente deprimidos y los que exhibieron síntomas graves de este trastorno. La tabla 1 puede entonces atomizarse en tres partes, como recoge la tabla 3.

Como permite apreciar este ejemplo hipotético, la asociación se ha «esfumado» dentro de cada uno de estos estratos, de modo que una vez «controlada» la depresión se pone en evidencia que el presunto efecto causal de la pérdida de peso era falaz. En esta ilustración el control de una sola variable condujo a descartar la relación causal aparente, pero podría ocurrir lo contrario: que tras el control de una variable confusora «emergiera» una asociación que en principio no se apreciaba. Por otra parte, lo típico es que sean varias las variables que se quieran controlar, y lo ideal es que dicho control se ejerza de manera simultánea**. El recurso de estratificar exige pérdida de información cuando la variable que se quiere controlar es cuantitativa (p. ej., edad) y por otra parte será casi seguramente inaplicable cuando son varias las variables que se quieren controlar a la misma vez, ya que ello demandaría formar quizá decenas de estratos y exigiría, por tanto, una muestra de miles de sujetos.

En la década de 1960 se empezó a desarrollar una técnica estadística bastante más compleja: la regresión logística. Su notable poder de síntesis, versatilidad y capacidad modeladora, una vez que se contó con rápidos ordenadores y avanzados programas (imprescindibles en este caso), convirtieron a esta técnica estadística en la más empleada de todas en la investigación biomédica contemporánea²³.

* En rigor se trata de un promedio ponderado, para el cual se puede computar un intervalo de confianza. No alambicamos este artículo con fórmulas que el lector puede hallar en la literatura estadística o de las que, simplemente, puede prescindir (si cuenta con un programa informático adecuado).

** Lamentablemente pudieran existir muchos otros factores de confusión, conocidos o no (p. ej., rasgos genéticos o muerte de uno de los padres).

TABLA 2. Cruzamiento de las variables intento suicida y pérdida de peso con presencia o ausencia de síntomas de depresión

	Deprimidos		Total
	Sí	No	
Casos	144	37	181
Controles	211	332	543
Total	355	369	724
PP	417	13	430
No PP	59	235	294
Total	476	248	724

PP: pérdida de peso.

TABLA 3. Distribución de casos y controles según grado de depresión y según hubiesen tenido o no una sensible pérdida de peso antes del intento suicida

	<i>No deprimidos</i>			<i>Moderadamente deprimidos</i>			<i>Muy deprimidos</i>		
	<i>PP</i>	<i>No PP</i>	<i>Total</i>	<i>PP</i>	<i>No PP</i>	<i>Total</i>	<i>PP</i>	<i>No PP</i>	<i>Total</i>
Casos	2	35	37	84	6	90	42	12	54
Controles	11	200	211	198	14	212	93	27	120
Total	13	235	248	282	20	302	135	39	174
$OR_1 = \frac{ab}{bc} = \frac{(2)(200)}{(35)(11)} = 1,0 \quad \quad \quad OR_2 = \frac{ab}{bc} = \frac{(84)(14)}{(6)(198)} = 1,0 \quad \quad \quad OR_3 = \frac{ab}{bc} = \frac{(42)(27)}{(12)(93)} = 1,0$									

PP: pérdida de peso.

Entre sus virtudes está el hecho de que suple por entero al enfoque de estratificación debido a Mantel y Haenszel en el sentido de que todo lo que éste puede potencialmente aportar queda como caso particular de lo que la regresión logística ofrece. Además supera al enfoque clásico, pues permite que con una muestra relativamente reducida el proceso de control arriba descrito se pueda verificar con varias variables a la vez, y, adicionalmente, lo hace independientemente del carácter de las variables que han de controlarse (éstas pueden ser dicotómicas, nominales, politómicas o continuas). Por ejemplo, en un estudio donde a mi juicio se emplea la metodología de casos y controles de manera ejemplar²⁴, debido al escrupuloso manejo de la temporalidad y a la cuidadosa selección de los 327 casos y 897 controles, se estudia la depresión como posible factor de riesgo de la enfermedad isquémica del corazón. Tras hallar que hay asociación apreciable entre este factor y la dolencia, emplean la regresión logística para corroborar que dicha asociación se mantiene luego de controlar simultáneamente hábito de fumar, diabetes, hipertensión y características del área donde vivía (socialmente deprimida o no).

3d. Sesgos

Como se comentó genéricamente en el subapartado 2d, son muchos los sesgos que pueden presentarse en cualquier tipo de estudio. En esta sección nos ocuparemos de dos expresiones que son típicas de los estudios de casos y controles: sesgo de memoria y sesgo de selección.

Para ilustrar el primero de ellos imaginemos que los casos son individuos con diagnóstico de enfermedad de Alzheimer y que se valora si el hecho de haber sido tratados con anestésicos generales pudiera constituir un factor de riesgo²⁵. Supongamos que para los casos la información se recaba de un familiar y para los controles del propio individuo. Es probable que cuando se trata de un enfermo el familiar haga un esfuerzo especial por recordar cualquier «exposición» que pudiera haberlo afectado potencialmente, en particular una operación que haya supuesto anestesia general. En cambio, un control

puede no tener la misma motivación para recordar, y es probable la información que comunique sea menos rigurosa. En tal caso se dice que la respuesta es diferencial entre los grupos, lo que podría introducir el susodicho sesgo de memoria. Probablemente en este ejemplo la asociación entre la exposición y el desenlace resultaría sobrestimada.

El sesgo más insidioso que puede producirse en los de estudios de casos y controles, sin embargo, es el de selección. Conseguir que no aparezca es, simplemente, la clave metodológica medular de los estudios de casos y controles. Los procedimientos para elegir los controles se pueden establecer adecuadamente siempre que se hayan fijado nítidamente los criterios empleados para seleccionar los casos. El grupo control teóricamente más apropiado corresponde a la subpoblación de individuos que, habiendo estado en riesgo de padecer el desenlace objeto de estudio, en el hipotético caso de haber llegado a él, habrían podido ser elegidos como casos.

Lo ideal sería que cualquier rasgo hubiese tenido en principio igual posibilidad para casos y para controles de estar presente al inicio del proceso que el estudio intenta reconstruir. Esto es, que ningún factor se halle más representado en uno de los grupos como consecuencia del modo en que los grupos fueron conformados. Dicho de otro modo: un factor puede aparecer con más frecuencia entre casos que entre controles (lo cual haría pensar que tiene un papel causal), pero hay que descartar que tal diferencia haya sido inducida por la forma en que se seleccionaron los grupos.

Esto quiere decir que los controles deben ser seleccionados de la misma cohorte imaginaria de la que proceden los casos. Cuando los casos proceden de una población bien delimitada en el tiempo y el espacio, la selección de controles se podría realizar mediante un muestreo aleatorio simple de dicha población. Siendo así, quedaría descartado sesgo alguno por este concepto. Sin embargo, esto se puede conseguir sólo raras veces.

En términos prácticos es crucial que los controles sean seleccionados independientemente de que hayan estado o no expuestos o no expuestos. Si la condición de exposición influye de alguna manera en la posibilidad de que un individuo sea o no incluido como control, se producirá un sesgo de selección.

Por ejemplo, supongamos que se valoran dos factores que han sido considerados de riesgo para la enfermedad de Alzheimer: consumo de carne cruda e hipotiroidismo. Supongamos que para estudiarlos se toman como casos todos los sujetos registrados como portadores de enfermedad de Alzheimer en el hospital de una región en la cual éste es el único servicio que atiende tal dolencia, y como controles, a una muestra de pacientes libres de la enfermedad que arriban al servicio de urgencia de ese hospital. En esta situación podríamos estar generando un sesgo de selección. En efecto, el consumo de carne cruda es más frecuente en zonas rurales que en el entorno urbano del hospital. Y ocurre que mientras el servicio que atiende el problema psiquiátrico recibe pacientes de toda la región, el servicio de urgencia sólo atiende casos urbanos (las urgencias de las zonas rurales acuden a centros de sus propios entornos). Siendo así, la muestra de controles tendría de antemano una subrepresentación de consumidores de carne cruda respecto de los casos, y ese monto adicional de consumidores de carne cruda, que nada tiene que ver con la enfermedad mental, podría contribuir a que erróneamente confirmemos peso causal en ella (ciertamente, el carácter causal de la relación entre carne cruda y la enfermedad ha sido descartada²⁶). En cambio, si el hipertiroidismo se presenta por igual en la zona atendida por el servicio de urgencia que en toda la región, el análisis concerniente a este factor no estaría sesgado por concepto de selección (la condición de factor de riesgo del hipertiroidismo es, en efecto, compatible con resultados hallados recientemente²⁷).

4. ADVERTENCIAS Y COMENTARIOS FINALES

Para concluir este trabajo examinaremos varios asuntos insuficientemente comprendidos y que son fuente de errores nada infrecuentes.

4a. Predicción y causalidad

No es difícil hallar trabajos publicados en los que tras hallar una OR alta se atribuya a una variable la condición de predictiva. Si cierto factor tiene peso causal en un desenlace, entonces siempre posee valor predictivo; e incluso puede tener tal valor un factor que no sea causal. Por lo general no hay incorrección en ello siempre que se haya contemplado la temporalidad (no necesariamente la plausibilidad, ni el efecto de terceros factores). Por ejemplo, la pérdida de peso puede tener valor predictivo a los efectos del suicidio aunque no tenga, como ya vimos, ninguna «responsabilidad» en el hecho. Y en esa calidad, puede tener interés a los efectos de la prevención. Sin embargo, lo que interesa especialmente enfatizar es que el recíproco es falso. «Factor de riesgo» y «factor predictivo» no son sinónimos: el hecho de que esté o no presente un factor de riesgo puede ocasionalmente ser útil para la predicción, pero una variable puede

hacer importante aportación a los efectos de predecir, aunque en sí misma no sea un factor de riesgo ni un componente real del entramado causal. Un factor puede servir para la predicción sin que eso lo convierta ni por asomo en un agente causal.

Otro asunto íntimamente relacionado con este tema merece alguna consideración. Debe quedar claro que con los datos de un estudio de casos y controles no se pueden estimar tasas de incidencia. Por ejemplo, sería simplemente disparatado concluir, a partir de la tabla 1, que el 28 % (128/450) de los que han perdido peso intentan suicidarse; o que esto ocurre con el 18 % (53/294) de los adolescentes que no han perdido peso. Tal imposibilidad es una consecuencia directa de que típicamente la muestra integrada por los casos y los controles seleccionados para el estudio no es representativa de la población. En efecto, si el desenlace que se estudia se presenta en la población con una prevalencia muy baja (p. ej., 2 por 1.000), el enfoque de casos y controles es el más adecuado*. De hecho, el rasgo más atractivo de los estudios de casos y controles consiste en que opera con muestras muy reducidas. De modo que en este ejemplo sería típico que se trabajara, digamos, con 200 casos y 400 controles: en tal caso, mientras que en la población hay un sujeto con la dolencia por cada 499 que no la padecen, en la muestra hay un enfermo por cada dos sujetos sanos.

Si se ha hecho un estudio de casos y controles y se ha identificado por su conducto una variable o condición de riesgo cabe preguntarse: ¿podrían, usando esa información, estimarse las tasas de incidencia? Es decir, ¿puede estimarse no sólo un valor aproximado del riesgo relativo (la OR), sino también la probabilidad de que un sujeto que tenga una condición de riesgo desarrolle la dolencia? La respuesta es afirmativa²⁸ siempre que se cuente con una estimación confiable de la prevalencia de la enfermedad (o desenlace) en la población. Esto es casi universalmente desconocido. En el anexo 1 se expone una solución del problema.

Cabe consignar que, por la misma razón de falta de representatividad, la función logística salida de un estudio de casos y controles no permite como tal estimar directamente la probabilidad de que se produzca cierto desenlace para un sujeto que tenga cierto perfil específico con potencialidad predictiva (sean las variables que integran el modelo causales o no). Sin embargo, es posible usar con ese fin la función logística si se hace una corrección de la estimación de uno de los parámetros suponiendo, nuevamente, que contamos con información auxiliar para ello²⁹.

4b. Tamaño de muestra

Es obvio que ningún análisis, independientemente del método elegido para llevarlo adelante, puede ser per-

*Un estudio prospectivo, por ejemplo, exigiría una muestra de muchos miles de sujetos para poder contar con estimaciones de tasas de incidencia mínimamente precisas.

ANEXO 1. Estimación de tasas de incidencia a partir de un estudio de casos y controles

Si llamamos p_1 y p_2 a las tasas de expuestos entre los casos y entre los controles, respectivamente, y f a la prevalencia general del desenlace en la población, entonces se puede demostrar por medio del teorema de Bayes que las tasas de desarrollo de la enfermedad (o aparición del desenlace) entre los expuestos y no expuestos, respectivamente, vienen dadas por las fórmulas siguientes:

$$t_1 = \frac{p_1 f}{p_1 f + p_2 (1-f)} \quad t_2 = \frac{(1-p_1) f}{(1-p_1) f + (1-p_2) (1-f)}$$

Por ejemplo, consideremos el caso de la depresión y los intentos suicidas resumido en la tabla 2. Si se admite que la tasa de intentos suicidas entre adolescentes de la comunidad es 2 por 1.000 ($f = 0,002$), recordando que

$$p_1 = \frac{144}{181} \text{ y } p_2 = \frac{211}{543}, \text{ se tendría:}$$

$t_1 = 0,00408$ y que $t_2 = 0,00067$. Recuérdese que en este ejemplo la OR es igual a 6,12; virtualmente igual a la razón de las tasas (RR) es

$$\frac{t_1}{t_2} = \frac{0,00408}{0,00067} = 6,09, \text{ como era de esperar.}$$

Lo mismo podría hacerse usando la tabla 1 para la pérdida de peso. Recuérdese que tiene sentido computar estas tasas independientemente de que se trate de un verdadero factor de riesgo o no; en definitiva, a los efectos de la prevención secundaria pudiera ser útil, por ejemplo, tener una estimación de la probabilidad de que un adolescente que ha bajado de peso intente suicidarse.

suasivo si no se cuenta con suficientes datos. Mucho se ha escrito acerca de cómo arribar a un tamaño de muestra mínimo, en general e incluso en relación con los estudios de psiquiatría³⁰. En realidad se ha escrito mucho, pero no muchas cosas diferentes; los artículos o capítulos de libros no pocas veces parecen clonados unos de otros. Existen asimismo numerosos programas informáticos diseñados para resolver este problema. EPIINFO 6.0 y EPIDAT 3.0 son, entre ellos, especialmente flexibles y amigables. No nos detendremos por tanto en dar pautas formales para encarar este importante problema. Considero oportuno, sin embargo, hacer una consideración muy general.

El problema del tamaño muestral ha sido históricamente manejado de una manera idílica: el discurso oficial de la estadística suprime la discusión sobre los notables escollos que supone aplicar las fórmulas y la suple con pautas operativas de muy dudosa aplicabilidad. Algunos autores son tajantes al afirmar que detrás de las fórmulas hay una falsa panacea. El afamado epidemiólogo Kenneth Rothman, por ejemplo, escribe³¹: «En resumen, el problema de determinar el tamaño de muestra más adecuado no es de naturaleza técnica, susceptible de ser resuelto por vía de los cómputos, sino que ha de

encararse mediante el juicio, la experiencia y la intuición.»

Pero, en general, esta realidad es ignorada o sepultada por el recetario que casi todos repiten anodinamente. Es útil comprender que cualquiera que sea el tamaño de la muestra, tanto los intervalos de confianza como los valores p , pueden ser calculados *a posteriori*. Consecuentemente, la sugerencia más general que puedo dar sobre este tema es que, sin desdeñar el valor orientativo que pueden tener los resultados surgidos de la aplicación de fórmulas, el tamaño muestral se elija sobre todo usando el sentido común, observando lo que se hace en la literatura y teniendo en cuenta los recursos de que se dispone. En el capítulo 11 de un libro publicado en 1997 doy cuenta pormenorizada de mis puntos de vista sobre este polémico asunto³².

Un problema conexo concierne a la relación que ha de fijarse entre número de casos y de controles. Lo más natural es que sean iguales. Aunque los controles suelen estar más disponibles que los casos, en muchas ocasiones la identificación e interrogatorio de los controles es un proceso largo y oneroso, hecho que puede aconsejar la igualdad. Sin embargo, tal equidad dista de ser necesaria, e incluso pudiera ser conveniente tomar más casos que controles. En efecto, cuando el número de casos es limitado, una alternativa natural es la de compensar de algún modo tal circunstancia incrementando el número de controles. El aumento en el número de controles por caso es útil hasta que se alcanza una razón de 4 por 1. De ahí en adelante no se gana mucho incrementando el tamaño de muestra a costa de aumentar el número de controles^{33,34}.

4c. Regresión paso a paso

Cualquier usuario de los paquetes estadísticos conoce la existencia de un recurso destinado a subseleccionar un modelo de regresión múltiple (es decir, un algoritmo que permite desechar algunas variables inicialmente consideradas para construir un «modelo final»). Lamentablemente, el cómodo recurso de la regresión paso a paso (*stepwise*), no sólo es estéril para el análisis de la causalidad, sino que puede ser simplemente pernicioso, por lo cual debe ser directamente eludido. Sólo tiene sentido cuando lo que se quiere es construir un modelo predictivo a partir de un estudio de casos y controles (lo cual exigiría hacer la corrección mencionada en el subapartado 4a).

A veces el procedimiento se emplea para descubrir cuáles son las variables causales y desechar por su conducto las que no los son. Otras veces se emplea, pero no queda nada claro su objetivo. En un estudio de casos y controles, para identificar los factores de riesgo para tornarse consumidor de antidepresivos³⁵ se hace una maniobra típica. Con las variables que exhibían una relación bivariable con el hecho de ser caso o control se ajustó una regresión logística a la que se aplicó luego el método *stepwise*. Según palabras de los investigadores,

tras aplicar este procedimiento «entraron a formar parte de la ecuación el hecho de haber vivido algún acontecimiento personal relevante y presentar valores de la escala de Zung* superiores a 50». Se ha construido un modelo «final»; ahora procede preguntarse ¿y qué? Daría la impresión de que se ha hecho de la aplicación de este recurso una finalidad en sí misma. Por lo menos en este trabajo concreto no se intenta sacar conclusiones (el hecho queda suelto y desconectado en el discurso), ya que cualquier respuesta sobre causalidad carecería de sentido. En efecto, sería imposible descartar que, al «quedar» la escala de Zung en el modelo, exista cierta variable asociada con dicha escala que resultó estadísticamente redundante (en virtud de lo cual, por ende, quedó fuera de la función obtenida con el *stepwise*), pero que pudiera tener un importante vínculo causal con el consumo de psicofármacos (máxime en este caso, en que lo que se «demuestra» a través de la regresión es, nada menos, que el consumo de antidepresivos asociado con estar deprimido). No es difícil hallar ejemplos de la literatura actual en que erróneamente sí se obtienen conclusiones³⁶.

El problema fundamental nace de la pretenciosa y a la vez ingenua interpretación que suele hacerse del resultado que arroja la regresión paso a paso. Su empleo con fines explicativos (identificar factores causales) es absurdo, pues la selección algorítmica de modelos no puede evitar que los resultados se deriven de meras concomitancias estadísticas (de hecho, en eso se basan), ni distinguir entre las asociaciones de índole causal y las debidas a terceros factores involucrados en el proceso. Consecuentemente, si bien el modelo de regresión logística puede ser de extraordinario interés para ayudar a entender los procesos biológicos y sociales luego de un estudio de casos y controles, como se hizo notar en el subapartado 3c, los procedimientos algorítmicos de subselección de variables para conformar un modelo «final» explicativo son totalmente improcedentes.

4d. Intervalos de confianza frente a pruebas de hipótesis

En la primera mitad del siglo xx los investigadores sanitarios, por lo general, no dominaban los métodos que permitieran cuantificar la evidencia y complementar así los razonamientos verbales de sus investigaciones. En ese lapso, sin embargo, lentamente se habían ido desarrollando y consolidando las llamadas pruebas de hipótesis. En efecto, en la década de 1920 había nacido un enfoque ideado por Ronald Fisher para medir el grado de incompatibilidad de los datos con una hipótesis y con él los famosos valores *p*; algunos años más tarde, Jerzy Neyman y Egon Pearson proponen un procedimiento cen-

trado en la elección entre dos hipótesis. Poco después empieza a gestarse anónimamente el recurso híbrido surgido de la fusión de ambos aportes³⁷. Retroalimentadas por la creciente universalización del acceso a poderosos recursos informáticos y el desarrollo de numerosos paquetes de *software* estadísticos (SPSS, SAS, BMDP, EPINFO, etc.) que ofrecen las pruebas de hipótesis como uno de sus principales atractivos, las pruebas de hipótesis consolidan su presencia casi universal en la investigación biomédica.

Sin embargo, paralelamente se fueron acumulando cuantiosas y persuasivas objeciones al uso de las pruebas de hipótesis. Los problemas señalados son de diversa naturaleza y no procede adentrarnos en ellos ahora, pero no quedan dudas de que son de enorme entidad y se han fundamentado a lo largo de más de 40 años, tanto en numerosos artículos científicos³⁸⁻⁴², como en libros de texto⁴³⁻⁴⁶. Para apreciar un resumen muy completo con casi 400 referencias que recorren los últimos 40 años, especialmente los últimos 10, véase el impresionante trabajo de Raymond Nickerson⁴⁷.

Como consecuencia, hace ya varios años que diversas revistas punteras de la producción científica internacional tienden crecientemente a no admitir trabajos en los cuales sólo aparezcan pruebas de este tipo. Por ejemplo, el *British Heart Journal* anunció en un editorial de 1988 que se unía a la demanda que ya desde 1986 había aparecido en el *British Medical Journal* cuando Gardner y Altman exhortaban a los autores a emplear intervalos de confianza en lugar de pruebas de significación. Tal postura pasó a ser compartida por revistas tan importantes como *The Lancet*, *Annals of Internal Medicine* y *American Journal of Public Health*⁴⁸ y, finalmente, fue adoptada por el Grupo de Vancouver⁴⁹ (Comité Internacional de Directores de Revistas Médicas), en cuyo apartado de requisitos técnicos dedicados al empleo de la estadística se consigna textualmente: «Se evitará la dependencia exclusiva de las pruebas estadísticas de verificación de hipótesis, tal como el uso de los valores *p*, que no aportan ninguna información cuantitativa importante.» Otros órganos normativos de la actividad científica se han ido poco a poco pronunciando en la misma dirección. (v., p. ej., el trabajo de Wilkinson⁵⁰ sobre las recomendaciones del Task Force on Statistical Inference creado por la Asociación de Psicólogos de Estados Unidos acerca del uso de la estadística en la investigación psicológica.)

Las debilidades intrínsecas del método, unidas a lo fácil que resulta interpretarlo erróneamente, al reclamo para que cese el uso adocenado que suele dársele a diario, y al espaldarazo que las graves objeciones han recibido de autoridades y comités como los citados, conducen a vaticinar que, más temprano que tarde, se consolide una nueva era donde el uso de los intervalos de confianza y la estadística bayesiana desplacen a las pruebas de hipótesis.

Aunque, a tenor de tal realidad cada día son más los autores y las revistas médicas que optan por eludir el empleo de este recurso, lo cierto es que la prueba de hi-

*La escala de Zung es un indicador propuesto en 1990 para la medición de la depresión, el cual se obtiene tras un cuestionario que el sujeto se autoaplica.

pótesis aún sigue siendo cotidiana moneda de cambio⁵¹, incluso en revistas formalmente adscritas al Grupo de Vancouver. Es obviamente razonable sugerir a los investigadores que se mantengan atentos a la evolución de este proceso y, de hecho, que se atengan a las normas explícitamente adoptadas por las revistas en que intenten publicar sus resultados.

4e. Lenguaje y conclusiones de un estudio de casos y controles

Tras encontrar una OR apreciablemente mayor que la unidad, muchos investigadores «concluyen» que las variables a la que corresponde dicha OR (o las que quedaron en el modelo si se ajustó una regresión logística, con *stepwise* o sin él) están asociadas con el fenómeno que se estudia (típicamente una enfermedad), eludiendo así el compromiso de pronunciarse acerca de si tales variables son o no causantes del fenómeno en cuestión. Del mismo modo que carece de sentido plantearse la asociación como un propósito (v. subapartado 2a), tampoco lo tiene convertir la constatación de que dos variables están asociadas en una conclusión, porque esta última debe ser el resultado de un proceso intelectual cualitativamente superior a la mera cuantificación fenomenológica que la primera representa y, en definitiva, porque las conclusiones deben ser respuestas a las preguntas que se formulan, y entre éstas no es legítimo preguntarse si hay o no asociación.

Por otra parte, el desconcierto terminológico existente es notable. En un artículo sobre el empleo contemporáneo de la regresión logística paso a paso³⁶ se comunicaba que al revisar trabajos publicados en la revista española *Medicina Clínica*, dentro del mismo texto y antecediendo a la expresión que alude a la enfermedad o problema que se estudia, hallamos las más diversas expresiones para calificar o denominar a las variables que son objeto de estudio. Aparte de que ocasionalmente esas expresiones son intrínsecamente cuestionables, el problema fundamental es que se emplean de manera aparentemente irreflexiva (incluso en un mismo artículo se les llama a veces de un modo y otras veces de otro a pesar de las diferentes connotaciones semánticas que tienen). Ocasionalmente se alude implícita o explícitamente a relaciones causales («factores que tienen un efecto sobre», «factores de riesgo», «variables influyentes en», «variables explicativas de», «factores más decisivos en»). Nada hay que objetar a estas expresiones si, en efecto, se trata de estudios en que se procura identificar factores explicativos del proceso estudiado. En otras ocasiones se alude a su potencialidad predictiva («predictores», «factores pronósticos de», «variables predictoras»), lo cual también puede ser correcto en ciertos contextos. Otras expresiones son directamente incorrectas; tal es el caso de «factores predictores del riesgo» (ya que no es el riesgo lo que se predice, sino un desenlace) o de «mediadores» (término totalmente indefinido y, por ende, ambiguo). Con frecuencia aparecen las consabidas alusiones a la asociación:

«variables asociadas con», «variables relacionadas con». Finalmente pueden hallarse giros, tales como «variables estadísticamente responsables de», algo a mi juicio directamente descabellado.

5. CONSIDERACIÓN FINAL

Como una última anotación me parece oportuno subrayar que la realización de un estudio de casos y controles metodológicamente sólido entraña un verdadero desafío que no puede resolverse siguiendo una receta; las ventajas prácticas que este enfoque exhibe respecto de la experimentación y de los estudios prospectivos observacionales son notables, pero demanda una alta dosis de reflexión crítica y supone siempre un ejercicio de creatividad. Quizás ello agregue para muchos un atractivo adicional, especialmente en un campo como el de la psiquiatría, en pleno desarrollo y con tantos problemas abiertos.

BIBLIOGRAFÍA

- Hill AB. Environment and disease: association or causation? *Proceedings of the Royal Society of Medicine* 1965; 58:295-300.
- Bofill J. Integración de los criterios de Sir Austin Bradford Hill a los criterios utilizados tradicionalmente para el estudio de la causalidad en valoración del daño corporal. 1999 (disponible en <http://www.la-plaza.com/vdc/BD/doc/revisiones/index.html>).
- Silva LC. La homeopatía, una crónica bicentennial. *Revista Habanera de Medicina* 2002;1(2). Disponible en http://www.sld.cu/instituciones/iscmh/rhabcm/rhbcm1/Articulos/articulo_luisc.htm.
- Davidson J, Gaylord S. Meeting of minds in psychiatry and homeopathy: an example in social phobia. *Alternatives Therapies in Health and Medicine* 1995;1(3):36-43.
- Carlston M. The mechanism of homeopathy. All that matters is that it works. *Alternat Ther Health Med* 1995;1: 95-6.
- Silva LC, Benavides A. Causalidad e inobservancia de la premisa de precedencia temporal en la investigación biomédica. *Rev Methodol (Belg)* 1999;7:1-11.
- Schulz KF, Grimes DA. Case-control studies: research in reverse. *Lancet* 2002;359:431-4.
- Schlesselman JJ. Case-Control studies. New York: Oxford University Press, 1982.
- Breslow N, Day NE. Statistical methods in cancer research Vol 1: The analysis of case-control studies 1980; IARC Scientific Publication No. 32 (Lyon: IARC).
- Szklo MF, Nieto J. Epidemiology: beyond the basics. Gaithersburg: Aspen, 2000.
- Cornfield J. A method of estimating comparative rates from clinical data. Application to cancer of the lung, breast and cervix. *J Nat Cancer Instit* 1951; 11:1269-75.
- Doll R, Hill AB. A study of the aetiology of carcinoma of the lung. *BMJ* 1952;2:1271-86.
- Silva LC. Diseño razonado de cuestionarios y muestras para la investigación sanitaria. Madrid: Díaz de Santos, 2000.
- Zornberg GL, Jick H. Antipsychotic drug use and risk of first-time idiopathic venous thromboembolism: a case-control study. *Lancet* 2000;356:1219-23.

15. Gardner MJ, Altman DG. Confidence intervals rather than p values: estimation rather than hypothesis testing. *BMJ* 1986;292:746-50.
16. Fairburn CG, et al. Risk factors for anorexia nervosa: three integrated case-control. *Comp Arch Gen Psychiatry* 1999;56:468-76.
17. McCreadie RG. Use of drugs, alcohol and tobacco by people with schizophrenia: case-control study. *Br J Psychiatry* 2002;181:321-5.
18. Vilalta J, Llinàs J, López S. Depresión y demencia: estudio caso-control. *Rev Neurol* 1999;29(7):599-603.
19. Kresnow M. An unmatched case-control study of nearly lethal suicide attempts in Houston, Texas: research methods and measurements. *Suicide and Life-Threatening Behavior* 2001;32(Suppl):7-20.
20. MacDonald S, et al. Nithsdale schizophrenia surveys 24: sexual dysfunction: case-control study. *Br J Psychiatry* 2003;182:50-6.
21. De Irala J, Martínez MA, Guillén F. ¿Qué es una variable de confusión? *Med Clin (Barc)* 2001;117:377-85.
22. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Nat Cancer Institute* 1959;22(4):719-48.
23. Silva LC, Pérez C, Cuellar I. Uso de métodos estadísticos en dos revistas médicas con alto factor de impacto. *Gaceta Sanitaria* 1995;9(48):189-95.
24. Hiipsley J, Fielding K, Pringle M. Depression as a risk factor for ischaemic heart disease in men: population based case-control study. *BMJ* 1998; 316:1714-9.
25. Broe GA, Henderson AS, McCusker E, et al. A case-control study of Alzheimer's disease in Australia. *Neurology* 1990; 40:1698-707.
26. Chandra V, et al. Case-control study of late onset «probable Alzheimer's disease». *Neurology* 1987;37:1295-300.
27. Breteler MMB, et al. Medical history and the risk of Alzheimer's disease: a collaborative re-analysis of case-control studies. *Internat J Epidemiol* 1991;20(Suppl 2):S36-42.
28. King G, Zeng L. Estimating Risk and rate levels, ratios, and differences in case-control studies. *Statistics in Medicine* 2002; 21:1409-27.
29. Silva LC. Excursión a la regresión logística. Madrid: Díaz de Santos, 1995.
30. Gómez C, Sánchez R. Cálculo del tamaño de la muestra en psiquiatría y salud mental (principios básicos para su estimación). *Rev Colomb Psiquiatr* 1998;27(2):131-42.
31. Rothman JK. *Modern epidemiology*. Boston: Little, 1986.
32. Silva LC. *Cultura estadística e investigaciones en el campo de la salud*. Madrid: Díaz de Santos, 1997.
33. Lasky T, Stolley PD. Selection of cases and controls. *Epidemiol Rev* 1994;16(1):6-17.
34. Gordis L. *Epidemiology*. Philadelphia: WB Saunders, 1996.
35. Rispau A. Factores de riesgo asociados al consumo de antidepresivos. *Atenc Prim* 1998;22(7):78-83.
36. Silva LC, Barroso I. Selección algorítmica de modelos en las aplicaciones biomédicas de la regresión múltiple. *Medic Clin* 2001;116:741-5.
37. Goodman SN. Toward evidence-based medical statistics (I): The p value fallacy. *An Intern Med* 1999;130:995-1004.
38. Edwards W, Lindman H, Savage LJ. Bayesian statistical inference for psychological research. *Psychologic Rev* 1963; 70:193.
39. Carver RP. The case against statistical significance testing. *Harvard Educ Rev* 1978;48:378-99.
40. Berger JO, Berry DA. Statistical analysis and the illusion of objectivity. *Am Sci* 1988;76:159-65.
41. Tamhane AC. Reseña sobre el libro Bechhofer RE, Santner TJ, Goldsman DM. *Design and analysis of experiments for statistical selection, screening and multiple comparisons*, John Wiley (New York), 1995. *Technometrics* 1996;38: 289-90.
42. Johnson DH. Hypothesis testing: statistics as pseudoscience. Fifth Annual Conference of the Wildlife Society, Buffalo, New York, 26 September, 1998.
43. Jeffreys H. *Theory of probability*, 3rd ed. Oxford: University Press, 1961.
44. Lindley DV. *Introduction to probability & statistics. Part 2: Inference*. Cambridge: University Press, 1970.
45. O'Hagan A. *Kendall's advanced theory of statistics. Vol 2B: Bayesian Inference*. London: Arnold, 1994.
46. Lee PM. *Bayesian statistics: an introduction*, 2nd ed. London: Arnold, 1997.
47. Nickerson RS. Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological Methods* 2000;5(2):241-301.
48. Evans SJW, Mills P, Dawson J. The end of the p value? *Br Heart J* 1988;60:177-80.
49. Comité Internacional de Directores de Revistas Médicas. *Requisitos uniformes de los manuscritos enviados a revistas biomédicas* (mayo 2001).
50. Wilkinson L. Task force on statistical inference. *Statistical methods in psychology journals*. *Am Psychologist* 1999; 54:594-604.
51. Sarria M, Silva LC. Empleo de las pruebas de hipótesis en la literatura biomédica iberoamericana actual. *Revista Cubana de Salud Pública*, 2003 (en prensa).