

Una Ceremonia Estadística para Identificar Factores de Riesgo

A Statistical Ceremony to Identify Risk Factors

Luis Carlos Silva Ayçaguer¹

¹Investigador Titular,
Universidad de Ciencias
Médicas de la Habana. Cuba.
lcsilva@infomed.sld.cu

RESUMEN Se exponen el contexto histórico y las condiciones en que se ha configurado lo que podría caracterizarse como una verdadera "cacería de factores de riesgo" en la epidemiología contemporánea. Entre otros problemas, se ha extendido la práctica de encarar el problema de la determinación o identificación de factores de riesgo a través de un ritual estadístico que roza la caricatura y, lo más importante, que la mayor parte de las veces no conduce a conseguir el objetivo de la investigación, ni a obtener nuevos conocimientos. El trabajo caracteriza dicho ritual y examina críticamente sus pasos. Se realiza una detallada discusión de la endeblez metodológica de tal procedimiento y se ilustra profusamente su empleo.

PALABRAS CLAVE Factores de Riesgo; Regresión Logística; Regresión paso a paso; Predicción; Factores de Confusión.

ABSTRACT What could be labeled as "risk factors hunt" has been taken place along the last years in epidemiological research. The historical context and conditions under which it has been developed are sketched on outline. Among other problems, a tangible tendency to face risk factors identification using a statistical ceremony has been taken place. It is pretty close to a caricature that, most of the time, is clearly deceptive, since does not lead us to achieve the research aims: to obtain new knowledge or a better understanding of health processes. Such a ritual is characterized and its steps are critically examined. An exhaustive discussion of their methodological fragility is exposed and its employment is thoroughly illustrated.

KEY WORDS Risk Factors; Logistic Regression; Stepwise Regression; Prediction; Confounders.

INTRODUCCIÓN

Por lo general, la investigación epidemiológica se emprende, bien con el fin de describir una realidad o proceso —es decir, para caracterizar en relativamente pocos trazos cómo es dicha realidad—, 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. El cumplimiento de la primera de estas encomiendas ha sido notable; pero el de la segunda se ha visto beneficiado por un puñado de éxitos y ensombrecido por recurrentes decepciones.

Hace unos años, en otro trabajo señalé (1), que ya resultaba imposible desconocer los inquietantes síntomas de crisis que vivía la epidemiología analítica observacional tal y como se ha venido practicando en los últimos lustros. En un incisivo e influyente trabajo publicado por la prestigiosa revista *Science*, Taubes (2) daba cuenta de ello a través de un discurso crítico vertebado a partir de las propias declaraciones, tomadas de aquí y allá, procedentes de figuras tan emblemáticas de la epidemiología actual como Greenland, Sackett, Rothman, McMahon, Breslow, Feinstein y Peto, donde los propios protagonistas iban aportando indicios de la desazón que prevalecía en sus predios metodológicos.

Expresado resumidamente, el síntoma esencial de tal crisis era claro e indiscutible: los resultados alcanzados por esta disciplina en la explicación de mecanismos causales han sido notoriamente modestos, máxime si se toman en consideración el tiempo y los recursos empleados; tras varias décadas de intensos y costosos estudios orientados a esclarecer la etiología de las más importantes dolencias que aquejan al mundo desarrollado, los dividendos son con frecuencia ficticios o decepcionantes. Como veremos, una parte importante de la responsabilidad de tales chascos ha sido el empleo de recursos estadísticos y sobre todo computacionales que se aplican de manera adocenada y acrítica. Desde entonces, la situación no ha hecho más que empeorar, acaso, paradójicamente, como consecuencia de la creciente facilidad para acceder a dichos recursos.

A partir de esta realidad, el debate relevante gira en torno a cuáles son los límites epistemológicos de una disciplina en la cual, como

en la meteorología o la demografía, generalmente no se puede acudir al recurso experimental y de la que podríamos quizás estar esperando más de lo que puede aportar.

La situación llegó por entonces a tal extremo que no era extraño hallar sombrías afirmaciones como las de Charlton (3) recogidas en las páginas del *Journal of Clinical Epidemiology*:

La epidemiología cada día se considera más a sí misma como una disciplina autónoma con sus propios patrones intelectuales para encontrar demostraciones (...) a pesar de que está sistemáticamente incapacitada para resolver debates concernientes a los mecanismos causales.

La realidad, afortunadamente, no es exactamente esa. A la vez que, ciertamente, se hacen más obvias y reconocidas las limitaciones de la epidemiología en su cauce actual, es injusto atribuir tal esterilidad a la disciplina como tal. Sus fracasos relativos no nos habilitan para descartar la posibilidad de que el enfoque epidemiológico pueda hacer contribuciones, aunque sean muchas veces parciales, a la explicación de los fenómenos, del mismo modo que sus éxitos históricos no pueden validar la sublimación de sus posibilidades, menos aun si no se trasciende el acomodamiento metodológico que hoy se observa.

Entre los problemas fundamentales que ponen de manifiesto una cultura epidemiológica deformada y reclaman enmienda se halla la sumisión al análisis que se verifica a nivel individual y el consecuente desprecio hacia el enfoque socioambiental, así como la sublimación y aplicación inercial de la estadística como sucedáneo y no como complemento de la reflexión crítica.

Vivimos en una época en que el sacrificio de "remilgos" humanistas en el altar de la eficiencia es frecuente moneda de cambio, y el "sálvese quien pueda" parecería ser la nueva filosofía de vida que se nos propone, y en no pocas ocasiones, se nos impone. En este marco general se consolida el enfoque reduccionista que deriva en una marcada sumisión inercial al paradigma que establece que las causas han de buscarse al mismo nivel en que se produce el problema. Las condiciones ambientales o contextuales se desdibujan casi por entero y se procede a la cacería

afanosa de factores individuales de riesgo (4). Los artículos que se orientan en esa dirección pueblan las revistas médicas como hongos después de la lluvia. Parece no repararse, por poner un ejemplo, en que alimentarse adecuadamente o no, si bien es una *acción* individual, está tan fuertemente condicionada por circunstancias económicas, publicitarias y sociales que no siempre constituye una *elección* individual.

Debe señalarse que la búsqueda de factores de riesgo para una entidad o desenlace dado no es en sí misma un esfuerzo censurable ni mucho menos. Cuando tal esfuerzo se verifica en un marco racional, teóricamente coherente y mediante métodos correctos, puede dar lugar a conocimientos útiles, aunque solo fuera para señalar avenidas de corroboración a las ciencias básicas, la experimentación o a nuevos y más refinados estudios observacionales. Lo impugnable radica en que con extrema frecuencia esta búsqueda se intenta desarrollar sin verdadera y genuina reflexión, apoyándose en la ingenua convicción de que, para que sea exitosa, basta emplear mecánicamente determinado algoritmo metodológico. En ese caso, el resultado puede ser contraproducente.

Y es ahí donde comparecen rituales estadísticos que crean una ilusión de eficacia que es peor que un espejismo. El presente artículo se orienta a discutir el más extendido de dichos rituales. Para conseguir una exposición más comprensible, en la Sección 1 se hace un breve repaso del concepto de causalidad y las condiciones básicas que han de cumplirse para demostrarla. A continuación (Sección 2) se describe el ceremonial estadístico que más extendidamente se oficia con tal finalidad y, de inmediato (Sección 3) se discuten las razones que fundamentan la naturaleza inconducente de dicho ritual. En la Sección 4, finalmente, expongo algunas consideraciones generales sobre el problema.

1. CAUSAS Y FACTORES DE RIESGO. EXIGENCIAS Y CONCEPTOS AFINES

La definición de qué debe entenderse por un *agente causal* y por un *factor de riesgo* es sin duda polémica. En mi opinión, cualquiera

que sea la definición formal de la primera categoría, *la definición funcional* –es decir, útil a los efectos prácticos de la prevención– de la causa de un efecto dado puede expresarse del modo siguiente: *cualquier factor, condición o característica, cuya supresión elimina la posibilidad de que se produzca el efecto, es una causa del mismo.*

Por otra parte están los elementos que, sin ser necesarios, son propiciatorios del efecto. Se trata de los llamados *factores de riesgo* (FR) para desarrollar una enfermedad (a), que son los que generan más confusión porque, como es lógico, suelen venir acompañados de las causas y "se parecen" a ellas, del mismo modo que las medidas paliativas para encarar una dolencia (por ejemplo, analgésicas) se parecen a la curación. Se trata de *factores susceptibles de ser modificados que, sin ser causas propiamente, pueden favorecer que el agente causal actúe.* Ejemplos típicos de factores de riesgo son la práctica de no usar anticonceptivos, circunstancia que favorece el embarazo y las enfermedades de transmisión sexual, el consumo de grasas saturadas a los efectos de las cardiopatías, o la obesidad como condición que propicia el surgimiento de dolencias vasculares.

Aunque, como se ha visto, se trata de categorías diferentes, a los efectos de lo que sigue es irrelevante si hablamos de una *causa* o de un *factor de riesgo*, ya que las exigencias metodológicas para la corroboración de uno u otro carácter en el marco observacional son básicamente las mismas. Sin embargo, sí procede subrayar que cuando se trata de estados que no pueden ser modificados mediante acciones o cambios de conducta (tal es el caso de la edad, el grupo sanguíneo o el género), no estamos ante un FR. Para dejar clara la distinción se les suele llamar *marcadores de riesgo.*

Son bien conocidas las listas de condiciones o premisas que han de cumplirse para tener derecho a considerar que cierta relación entre dos variables es causal o que la presencia de un factor entraña efectivamente riesgo a los efectos de enfermar. Aunque hay requisitos obvios, tales como la *plausibilidad biológica* (premisa que exige que la hipótesis de causalidad cuente con cierto respaldo teórico), aquellas verdaderamente cruciales y que han de ser insoslayablemente atendidas desde el punto de vista metodológico son las siguientes:

1. *congruencia temporal*: la presunta causa ha de haber actuado antes que el efecto y el diseño del estudio tiene que garantizar que las mediciones se hayan realizado con acuerdo a ello;
2. *ausencia de sesgos*: han de evitarse circunstancias que conduzcan a que veamos distorsionadamente en un sentido u otro la realidad;
3. *asociación*: debe corroborarse que la presunta causa o FR debe estar asociada con la variable que registra el desenlace (enfermedad);
4. *carácter intrínseco de la asociación observada*: hay que corroborar que la asociación no puede ser enteramente atribuible a la acción de otros factores; es decir, hay que constatar que la asociación subsiste una vez "controlados" los llamados factores de confusión, que en esencia son variables que influyen sobre las dos que se examinan de manera que pueden crear la ilusión de que estas últimas guardan una relación intrínseca (b).

Para garantizar el cumplimiento de los dos primeros requisitos, el esfuerzo ha de concentrarse en el empleo de diseños adecuados. La corroboración de los últimos dos, sin embargo, se puede conseguir básicamente mediante procedimientos estadísticos. Coherentemente con el propósito del presente trabajo, habré de ceñirme a dichos procedimientos.

1.1 ASOCIACIÓN

La asociación es condición *sine qua non* de la causalidad y puede medirse a través de diversos estadígrafos tales como, por ejemplo, un riesgo relativo, un *odds ratio*, un coeficiente de correlación o una diferencia de medias aritméticas. Cuanto más intensa sea, más verosimilitud adquiere la hipótesis de que hay una relación de causalidad directa o indirecta. Ello se debe a que, tras identificar una asociación en los estudios observacionales, una de las más importantes incertidumbres estriba en que la correlación observada pudiera ser solo un reflejo del efecto que ejercen otros condicionantes; si la correlación es intensa, la posibilidad de que la asociación pueda explicarse enteramente por el efecto de tales condicionantes, los llamados factores confusores, es mucho menor.

1.2 CONTROL DE FACTORES CONFUSORES

Los diseños 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 son conflictos potenciales que quedan conjurados por el propio diseño. Sin embargo, la virtud fundamental consiste en que a través de la asignación aleatoria, especialmente si las muestras son grandes, se equipara o balancea a los grupos que van a compararse respecto de todos los factores que pudieran ser confusores, conocidos o no.

Lamentablemente, imperativos éticos y prácticos de índole diversa suelen imposibilitar la experimentación en el marco de la epidemiología. Los estudios observacionales de cohorte son los sucedáneos naturales 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 de la enfermedad entre dichos grupos. De ese modo, 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. Por eso suelen preferirse los *estudios de casos y controles*, a pesar de que ellos pueden ofrecer dificultades para asegurar el cumplimiento de la premisa de precedencia temporal (6).

Ahora bien, el gran problema que afecta a todos los estudios no experimentales lo ofrece el control de factores confusores. ¿Cómo asegurarse de que la asociación observada es intrínsecamente causal y que no se debe exclusivamente a la influencia de al menos un factor (quizás varios factores) sobre los dos que se examinan? Como se ha expuesto en cientos de artículos y decenas de libros, el recurso clásico para ello era la llamada *postestratificación* y posterior aplicación de las técnicas propuestas por Mantel y Haenszel a finales de la década de los 50 del siglo pasado (7). Sin embargo, tal recurso está virtualmente ausente en la actualidad de las aplicaciones reales por su naturaleza restringida y artesanal, y ha quedado casi por entero relegado a la nostálgica exposición que de él se hace en los cursos ortodoxos de epidemiología. Su lugar ha sido ocupado por los modelos multivariados, muy especialmente por la *regresión logística*.

1.3 LOS MODELOS MULTIVARIADOS

En efecto, los modelos multivariados (o multivariantes) y, destacadamente, la regresión logística, constituyen una alternativa sumamente atractiva y útil a los efectos de controlar simultáneamente de manera ágil y elegante muchos posibles factores de confusión. A su popularidad contribuye decisivamente el hecho de que las enormes dificultades computacionales que ofrece debido a la complejidad de los algoritmos en que se basa han dejado de ser un obstáculo con el advenimiento de las rapidísimas computadoras personales y los numerosos programas informáticos a los que hoy todos tenemos acceso.

En esencia, si consideramos la respuesta binaria "enfermar-no enfermar" y ajustamos una regresión logística introduciendo una posible causa o FR como variable independiente y un conjunto de covariables con vistas a ser controladas, se obtiene un coeficiente para la primera que permite estimar el grado de asociación entre dicha posible causa y la respuesta "ajustado por el posible efecto confusor de todas las covariables incorporadas al modelo". Bien usada, esta propiedad constituye un excelente paliativo a la limitación que supone el hecho de no poder asignar aleatoriamente los sujetos a los tratamientos.

1.4 EXPLICACIÓN Y PREDICCIÓN

En este marco debe advertirse algo recurrentemente olvidado: "factor de riesgo" y "factor predictivo" no son sinónimos: el hecho de que esté actuando o no un factor de riesgo (y conocer el grado en que éste gravita) 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 tenga papel causal alguno. Por ejemplo, la pérdida sostenida de peso puede tener valor predictivo a los efectos de que un adolescente realice un intento suicida aunque la pérdida de peso no tenga –como es obvio– ninguna "responsabilidad" en el hecho. Y en esa calidad, puede tener interés a los efectos de la prevención. Pero lo que interesa especialmente enfatizar es que el recíproco es falso. Un factor puede servir para la predicción sin que eso lo convierta ni por asomo en un agente causal. Por ejemplo, los marcadores de riesgo suelen tener una valor predictivo, pero como ya vimos no pueden considerarse factores de riesgo.

Consecuentemente, resulta absurdo que en un trabajo concebido para identificar FR se concluya (o se afirme en la discusión) que este o aquel factor es "predictivo". Por lo general no hay incorrección en ello siempre que, además de la asociación, se haya contemplado la temporalidad (no necesariamente la plausibilidad, ni el efecto de terceros factores). Pero es crucial comprender que también puede tener tal valor un factor que no sea causal.

2. EL ESQUEMA DE UNA CEREMONIA METODOLÓGICA

Las consideraciones recordadas de manera sintética en la sección precedente nos colocan en el escenario donde se despliega la ceremonia estadística que será objeto de análisis en este trabajo.

Un procedimiento sumamente extendido hasta el punto de haberse convertido en un verdadero ritual, y que se aplica intensamente en estudios observacionales orientados a identificar factores de riesgo para desarrollar o padecer cierta

enfermedad (estudios de casos y controles y estudios prospectivos) discurre básicamente del modo siguiente:

1. Identificar un conjunto de variables que verosímilmente pudieran considerarse causas o factores de riesgo; para ello se acude a criterios de expertos, estudios previos, consideraciones teóricas o búsquedas en manuales o en Internet.
2. Hacer un estudio "univariado" donde se valore asociación entre cada una de las variables arriba mencionadas con la variable de respuesta (típicamente una enfermedad o una complicación) mediante una prueba de significación (por lo general Ji-Cuadrado o t de Student en dependencia de que la variable sea dicotómica o continua respectivamente).
3. Ajustar un modelo multivariado con las variables originalmente identificadas. Lo más frecuente es que se aplique la regresión logística y que el ajuste se realice con aquellas variables independientes para las que se obtuvo "significación" en el paso anterior.
4. Identificar las variables independientes que "se asocian" con la variable de respuesta. En este punto la receta tiene dos variantes básicas:
 - 4a. identificar cuáles de las variables mantienen su carácter significativo tras haber hecho el ajuste mencionado en el punto 3;
 - 4b. aplicar un procedimiento algorítmico de selección de variables; es decir, aplicar una selección "paso a paso" (*stepwise regression*) para determinar cuáles variables han de "quedarse" en calidad de factores detectados como verdaderamente influyentes y cuáles habrán de despreciarse.
5. Calificar como factores de riesgo a las variables identificadas en 4a. o a las que se hayan conservado en el "modelo final" si se hubiera apelado a la regresión "paso a paso".

Debe puntualizarse que el ritual descrito no se emplea siempre de manera exactamente

igual. En ocasiones, por ejemplo, se omite el paso 2 y se procede del primer paso directamente al 3. A veces, al cumplir el paso 5 no se habla de "factores de riesgo" como tal sino que se emplea una expresión neutra: "tales y más cuales son las variables que se asocian" con la dolencia. Tal declaración es especialmente inútil e intrascendente, pues la asociación *per se* nunca interesa; solo tiene interés en tanto ingrediente de un proceso (8).

Hallar ejemplos en que se aplique este esquema es sin duda más fácil que encontrar estudios que se planteen resolver la tarea de la causalidad y que prescindan de ellos. Una rapidísima búsqueda en Internet, solo en idioma castellano, me permitió hallar en menos de una hora decenas de ejemplos en diversas revistas latinoamericanas y españolas. La situación en el mundo anglosajón no es muy diferente.

Solo a modo de ejemplo, se invita al lector a revisar una reducida muestra de ejemplos recientes en Venezuela, Chile, Perú, Colombia, España, México, Costa Rica, Argentina y Cuba, todos susceptibles de ser hallados en Internet (9-26).

3. LA ESTERILIDAD DE LA CEREMONIA

A continuación desarrollo un sucinto análisis orientado a demostrar que la aplicación canónica de este procedimiento es inconducente. Veamos cada uno de los pasos.

→ Identificar un conjunto de posibles factores de riesgo

En primer lugar, hay que señalar que una tendencia bastante generalizada es la de plantearse la tarea de identificación de "todos" los factores de riesgo para una dolencia. Curiosamente, este ampuloso propósito (aunque no tiene nada objetable en principio) es mucho más típico de ambientes de menor desarrollo que de espacios académicos consagrados. Dicho de otro modo: cuanto más desarrollo, más incisivas, puntuales y creativas son las preguntas; cuanto menos desarrollo, más abarcadores y trillados son los propósitos.

El otro asunto es que al trazarse este propósito global, los investigadores suelen poner a prueba ciertos factores cuyo carácter de riesgo o cuyo papel causal es bien conocido. Si ellos se hubieran planteado algunas preguntas puntuales, obviamente no destinarían energía a examinar condiciones o variables para las que ya se tiene una respuesta hace mucho tiempo.

→Valorar significación a través de un estudio "univariado" con cada variable

Suele considerarse que un riesgo relativo o un *odds ratio* mayor que 2 es suficientemente grande como para considerarse promisorio o como para prestarle atención (27). Como se consigna en el citado artículo de Taubes (2), algunas de las importantes figuras de la epidemiología contemporánea demandan valores mayores para dignarse a prestarle atención; Marcia Angell, por ejemplo, ex editora de *New England Journal of Medicine* reclamaba que éste ascendiera a por lo menos 3. Lo que sí no menciona ningún experto como criterio demarcatorio es que la asociación hallada difiera significativamente de la nulidad.

Una razón para tal omisión reside en la bien conocida circunstancia de que lo que realmente importa es la magnitud del efecto (el grado de la asociación) y que el carácter significativo es irrelevante, habida cuenta de que la famosa "significación" puede conseguirse tomando muestras suficientemente grandes. Dicho de otro modo, cuáles de las variables elegidas van a ser significativas es algo que si bien depende de la magnitud de la asociación, también depende vitalmente de un elemento exógeno a la realidad: cuántos recursos se tengan y, por tanto, cuán grande sea el tamaño muestral (28).

La socorrida y universal *prueba de significación estadística* (PSE) ha acumulado cuantiosas y persuasivas objeciones a lo largo de los últimos decenios. Para apreciar un resumen muy completo con casi 400 referencias que recorren los últimos cuarenta años del debate, especialmente los últimos diez, véase el impresionante trabajo de Nickerson (29). No procede en este artículo extendernos sobre un asunto que puede consultarse en numerosos lugares.

Pero vale la pena reparar en que hace muchos años que diversas revistas punteras de la

producción científica internacional tienden crecientemente a no admitir trabajos en los cuales los análisis dependan de las pruebas de este tipo. Ya desde 1986 había aparecido en *British Medical Journal* la demanda de prescindir de las PSE y se exhortaba 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 *Lancet*, *Annals of Internal Medicine* y *American Journal of Public Health* y, finalmente, fue adoptada por el *Grupo de Vancouver* (30), 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 sobre la magnitud del efecto". Otros órganos normativos de la actividad científica se han ido poco a poco pronunciando en la misma dirección (c).

Aunque, a tenor con 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 hipótesis aún sigue siendo cotidiana moneda de cambio (32), incluso en revistas formalmente adscriptas al *Grupo de Vancouver*. Si bien las PSE son merecedoras, como se ve, de una valoración crítica por sí mismas, el asunto es mucho más grave cuando se comete la ingenuidad de emplearlas como un cribado inicial para la identificación posterior de causas o de FR.

→Identificar cuáles de las variables mantienen su carácter significativo tras ajustar un modelo multivariado

El ajuste de un modelo multivariado para controlar variables confusoras es en principio adecuado. Sin embargo, la manera en que casi siempre se emplea es desastrosa. Lo correcto sería lo siguiente:

- identificar con claridad la variable cuya condición de FR (o posible causa) se quiere evaluar e incluirla como una de las variables independientes del modelo;
- valorar teóricamente uno por uno aquellos factores que pudieran ser candidatos a ser confusores

- (para ello hay que corroborar que cumple las condiciones) e incluir los que estén en ese caso como covariables del modelo;
- valorar la magnitud de la asociación que se obtiene luego de haber controlado todos los factores del punto anterior;
- valorar esta magnitud a la luz de lo que se obtuvo cuando se hizo la estimación cruda (sin ajustar por los factores confusores).

Nótese que según lo anterior, el control de variables en el contexto observacional no debe practicarse si no es para evaluar hipótesis claramente expuestas y fundamentadas con anticipación, donde el potencial papel confusor de las variables que se controlan también tenga respaldo teórico independiente y fundamento lógico anterior.

En cambio, lo que suele hacerse, simplemente, es:

- incluir en el modelo todas las variables que se consideran posibles factores de riesgo (háyase o no realizado el cribado univariado);
- identificar todas aquellas cuyos coeficientes de regresión sean significativamente diferentes de cero.

Al depositar en este algoritmo la responsabilidad de decidir cuáles son los FR se pasan por alto al menos los siguientes conflictos:

- a) las ya mencionadas limitaciones de las PSE;
- b) que probablemente se estén controlando factores que no son confusores (un FR no necesariamente es un confusor para la asociación de la enfermedad con otro FR);
- c) la frecuente inclusión de marcadores de riesgo como si fueran necesariamente confusores.

→Concluir que las variables influyentes son las que se retienen tras aplicar una selección "paso a paso"

Consideremos ahora los *métodos de selección algorítmica del modelo*, que constituyen el caso extremo de la ingenuidad cuando se aplican en el contexto del ceremonial descrito. Estos métodos se han concebido para identificar aquellas variables que habrán de integrar la fun-

ción que a la postre será empleada como *modelo resumen* del proceso bajo estudio. No explicaré aquí en detalle cómo operan estos métodos; basta comprender que la lógica subyacente de tal recurso consiste en conservar las variables independientes que contienen información relevante y a la vez prescindir de aquellas que resulten redundantes respecto de las que quedaron en el modelo. Se trata de procedimientos de índole exclusivamente estadística, que discurren según algoritmos programables en los que, una vez elegido el conjunto inicial de variables, no intervienen los juicios teóricos de los investigadores.

Se han ideado varias alternativas para seleccionar un "modelo resumen". La más conocida es la llamada *regresión paso a paso* -RPP- (*stepwise method*), susceptible de ser aplicada según diversas variantes. Las más empleadas son: ir incorporando variables al modelo (*forward selection*), e ir eliminando variables de él (*backward elimination*). Virtualmente todos los grandes paquetes informáticos para el tratamiento estadístico de datos (tales como SPSS, SAS, BMDP o MINITAB) brindan la posibilidad de aplicar estas dos posibilidades.

Supongamos que un investigador quiere construir una función que permita estimar cuán probable es que una embarazada dé a luz un niño con bajo peso atendiendo a un conjunto de rasgos de la madre tales como escolaridad, si fumaba durante el embarazo, edad, número de hijos anteriores, tensión arterial, salario, etc. En tal caso, el modelo de RL (incluso con el colofón de una RPP) puede ser útil. Es posible que en el modelo final no figure el hábito de fumar (que tiene responsabilidad causal) y sí el salario (que no la tiene); sin embargo, esto no ha de preocupar, ya que la información inherente al hábito nicotínico ya esté subsumida en otras tales como el propio salario y la escolaridad (ya que fuman más las madres de menor nivel socioeconómico). Es decir, si este procedimiento se emplea para construir un modelo predictivo a través de un proceso acorde con el llamado "Principio de Parsimonia" (tratar de que figure la menor cantidad posible de variables), no veo objeción de peso al empleo de la RPP.

Sin embargo, por lo general (aunque ocasionalmente se use la terminología de la predicción) la RPP se emplea para identificar los factores

verdaderamente influyentes y no para construir modelos de predicción. De hecho, autores de gran prestigio sugieren que se ajuste un modelo de regresión múltiple y de inmediato se aplique un procedimiento algorítmico (33).

Sus resultados suelen ser interpretados como sigue: las variables que se "quedan" dentro del modelo final son parcialmente responsables (y quizá las principales y hasta las únicas responsables) de las modificaciones que experimenta la variable dependiente; las que no permanecen, o bien no influyen causalmente en el proceso, o su influencia no es apreciable. Es decir, muchos investigadores utilizan la selección algorítmica de modelos con la aspiración de obtener de manera automática conclusiones explicativas sobre el proceso causal que estudian.

La esperanza de que el uso de estos procedimientos contribuya a "entender" o "explicar" la realidad es, en el mejor de los casos, estéril o quimérica; y, no con baja probabilidad, contra-productiva y descabellada.

Para interpretar más claramente esta realidad, consideremos un ejemplo expuesto en Silva y Barroso (34). Supongamos que un investigador estuviera interesado en conocer cuáles son los factores que verdaderamente influyen en el padecimiento de enfisema pulmonar antes de los 60 años de edad. Para ello podría comenzar seleccionando un conjunto de variables presuntamente explicativas del proceso que se estudia, tales como edad, antecedentes patológicos familiares, hábito de fumar cigarrillos, pigmentación de los dedos (pulgar, índice y medio) de la mano, polución ambiental en su medio laboral, etc. Podría entonces realizar un estudio prospectivo, aplicar la regresión logística y, finalmente, realizar una selección mediante RPP con la idea de establecer cuáles de estas variables tienen mayor peso causal sobre la aparición de enfisema (o cuáles los factores que entrañan mayor riesgo de desarrollar la dolencia).

No sería para nada inverosímil que tal acción tuviera como posible resultado que el grado de pigmentación en los dedos quedara incluido en el modelo y que no ocurriera lo mismo con la condición de fumador, debido a la madeja de correlaciones que tienen estas dos variables con las restantes del modelo y a la asociación que tienen entre sí. Si no fuera por la cer-

teza que existe hoy de que el hábito de fumar es una práctica que favorece el enfisema pulmonar y de que la pigmentación de los dedos (inducida por el hábito) no guardan una relación causal (d), el investigador sería conducido por sus propias reglas de análisis a aseverar que la pigmentación de los dedos, en caso de que quedase incluida en el modelo final, es un factor de riesgo para esta entidad. Tal conclusión, lejos de iluminar el camino hacia el conocimiento de las verdaderas relaciones causales, lo ensombrecería o, más bien, lo obstaculizaría. Conviene no olvidar que "los números no saben de donde vienen", como advirtiera Lord (35) hace más de cincuenta años.

Si el modelo de regresión se aplicara para estimar la probabilidad de que un sujeto con cierto perfil desarrolle en el futuro un enfisema pulmonar (por ejemplo, para emprender una especial acción preventiva sobre quienes tengan tal perfil), entonces la pigmentación de los dedos podría figurar con todo derecho en el modelo. En efecto, en este caso el enfoque ha de ser pragmático: si se consiguen buenas predicciones, poco importan los medios. En este contexto resulta irrelevante si una variable dada ha quedado incluida en virtud de que desempeña un papel causal, o por ser un mero reflejo de otras variables que no figuran pero que sí pudieran tener influencia. Todo lo que importa en ese caso es construir un instrumento eficiente para la predicción. Pero si se quiere entender el mecanismo de producción del enfisema (identificar factores causales o de riesgo), al calificar de "predictivas" a ciertas variables independientes no se está avanzando absolutamente nada pues, aunque lo fueran, se está respondiendo a la pregunta equivocada.

Lo curioso es que, a la vez que muchos autores están avizorados acerca de que no deben confundir asociación con causalidad en el marco univariado, parecen olvidarlo cuando quedan encandilados por los métodos multivariados así como que no comprendan que, al aplicar estos algoritmos mecánicamente, están incurriendo solapadamente en el viejo sofisma. Variables que pudieran tener responsabilidad "directamente causal" pueden resultar eliminadas al ser suplidas por una o más variables que no tengan influencia real alguna pero que se vinculen con ella. Y en la medida que el asunto se dirime en la caja negra de la RPP, nada podemos hacer para evitarlo.

Por otra parte, cabe recordar una vez más que estos criterios están asentados sobre las pruebas de significación; por lo tanto, su pertinencia está sujeta a todas las suspicacias que ellas despiertan, como ya vimos arriba.

Hace casi tres decenios, Guttman (36) había advertido que "el uso de la regresión paso a paso es en la actualidad una confesión de ignorancia teórica sobre la estructura de la matriz de correlaciones". Cuando la regresión múltiple se usa para describir los patrones de causalidad según los cuales ciertas variables actúan sobre otra, la regresión paso a paso equivale a cubrir esa ignorancia con un algoritmo que piense por el investigador. No en balde el *stepwise regression* fue rebautizado irónicamente por Leamer (37) como *unwise regression* (e).

Ocasionalmente, el ceremonial descrito ni siquiera se completa: se construye un modelo "final", pero todo se queda en eso. En tales casos daría la impresión de que se ha hecho de la aplicación de este recurso una finalidad en sí misma.

Veámoslo en un estudio real de casos y controles concebido para –según se deriva del propio título del estudio– identificar factores de riesgo para que un sujeto se torne consumidor de antidepresivos. Allí Rispaú A *et al.* (38) realizan la maniobra típica pero sin realizar siquiera un cierre conceptual. Con las variables que exhibían relación significativa a nivel bivariante 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 el presentar valores de la escala de Zung (f) superiores a 50". Se ha construido un modelo "final"; ahora procede preguntarse ¿y qué? 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 haya quedado fuera de la función obtenida con el *stepwise*) pero que pudiera tener 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 está asociado con estar deprimido).

En síntesis, el problema fundamental está en la pretenciosa y a la vez ingenua interpretación que suele hacerse del resultado que arroja la RPP. Su empleo con fines explicativos 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 los modelos de regresión múltiple pueden ser –en caso de que se hayan aplicado razonadamente– de extraordinario interés para ayudar a entender los procesos biológicos y sociales, los procedimientos algorítmicos de subselección de variables para conformar un modelo "final" explicativo son, salvo situaciones excepcionales, totalmente improcedentes.

4. CONSIDERACIONES FINALES

En este punto lo primero que cabe hacer es una apelación al sentido común: si el empleo mecánico de recursos estadísticos, sean univariados o multivariados, pudiera ayudar a esclarecer las complejas relaciones causales que expliquen por qué unos individuos enferman y otros no, entonces con las enormes bases de datos hoy disponibles, las poderosísimas y veloces computadoras actuales y los potentes programas informáticos a los que todos tenemos acceso, la etiología del cáncer de mama, por poner un ejemplo, no sería el misterio que es hoy para la ciencia y que obliga a la generación apremiante de nuevos enfoques que se han venido reclamando (39), muy alejados de la acomodaticia esperanza de que el software puede suplir nuestras perspicacia y creatividad.

Se ha argüido que las enfermedades, en especial las coronarias y los tumores malignos, son entidades muy complicadas y dependientes de demasiadas variables mutuamente correlacionadas para que el análisis pueda prosperar con recursos estadísticos elementales. Tal afirmación es un poco ridícula, no porque sea falsa, sino porque es

verdadera para casi cualquier dolencia imaginable (desde el catarro común hasta la esclerosis múltiple). Pero en cualquier caso, lo que cabe subrayar es que, a partir de esta convicción, en los últimos años se ha multiplicado el empleo de complejos modelos estadísticos multivariados que supuestamente podrían esclarecer las cosas. La ocasional esterilidad del control de factores resulta en este caso menos obvia, y muchos usuarios quedan engeguados por las luces deslumbrantes del análisis multivariado. Pero la situación es exactamente la misma: el problema no radica en la frondosidad del método estadístico sino en la manera errónea de emplearlo y en la falta de teorización. De hecho, se dispone hace ya mucho de avanzados recursos tanto estadísticos como computacionales; esto no parece haber producido, sin embargo, un giro visible en la situación.

No casualmente Kolmogorov (40), uno de los más eminentes especialistas en teoría de probabilidades, advertía:

...en el estudio de fenómenos tan complicados como los que aparecen en biología y sociología, los métodos matemáticos no pueden desempeñar el mismo papel que, por ejemplo, en la física. En todos los casos, pero especialmente allí donde los fenómenos son más complicados... si no queremos perder el tiempo manejando fórmulas desprovistas de significado, el empleo de la matemática es útil sólo si se aplica a fenómenos concretos que ya han sido objeto de una profunda teorización.

La aplicación de la RPP es la situación en que el ritual se lleva a su máxima expresión.

Se ha enfatizado la necesidad de incluir solo variables cuyo sentido epidemiológico o clínico esté claro, pero eso no resuelve el problema cardinal: si aspiramos a que la RPP nos conduzca a conseguir conocimientos que no teníamos sobre la preeminencia causal de unas variables sobre otras, necesariamente tendremos que incluir variables iniciales cuyo papel ignoramos; y viceversa: si conocemos cabalmente ese papel, entonces no puede decirnos nada novedoso, y la selección algorítmica del modelo se convierte en una finalidad y no en un medio. No es posible escapar de este laberinto.

Los esfuerzos por resolver el problema a través de poderosos programas informáticos recuerdan a los constructores de máquinas de movimiento perpetuo: ignorantes de la ley de conservación de la energía y creyendo que sus fracasos se debían a que el diseño del aparato no era suficientemente ingenioso, procedían a desgastarse en la confección de nuevos y más sofisticados dispositivos (41) (g).

Nótese finalmente que, bajo ningún concepto, se está desdenando el poder esclarecedor de la estadística; lo que se rechaza es la extendida ilusión de que ella pueda aportar explicaciones por sí sola. Lo que se quiere subrayar es que resulta más eficiente no perder el tiempo buscando o aplicando códigos algorítmicos de procedimiento y recetas estadísticas para hacer una epidemiología fructuosa. El pensamiento científico nunca ha sido algoritmizable en ningún dominio. ¿Por qué habría de serlo en éste? La estadística puede ocupar un espacio como invaluable recurso instrumental, pero solo un régimen de amplia simbiosis transdisciplinaria y metodológica es capaz de cerrar ciclos cognoscitivos integrales.

NOTAS FINALES

a. Aunque por lo general el "efecto" de que hablamos es una enfermedad, no siempre es ese el caso. Pudiera tratarse de la realización de un intento suicida, de una complicación dentro de un proceso morboso, de sufrir un accidente o de cualquier otro desenlace, adverso o no. No obstante, en lo sucesivo aludiremos regularmente a "enfermedad", que es la situación más frecuente en el marco de los factores de riesgo.

b. Una definición formal con su correspondiente discusión y varias ilustraciones puede hallarse en de Irala, Martínez y Guillén (5).

c. Véase por ejemplo el trabajo de Wilkinson (31) 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.

d. Es obvio que si se inventara un jabón que borrara todo pigmento, ello no mejoraría las perspectivas de un fumador en cuanto al desarrollo de un enfisema.

e. Juego de palabras intraducible que aprovecha que el vocablo *wise* denota en inglés la manera o el modo de hacer algo, pero también significa "sabio", de modo que *unwise regression* vendría a ser algo así como regresión tonta o irracional

f. 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 auto-aplica.

g. Incidentalmente, no deja de ser asombroso que tales inventores no se den por vencido ni siquiera en pleno siglo XXI, y no sin éxito financiero y mediático, como denuncia Park (42).

REFERENCIAS BIBLIOGRÁFICAS

1. Silva LC. Hacia una cultura epidemiológica revitalizada. *Revista Humana*. 1997;1(5):23-33.
2. Taubes G. *Epidemiology faces its limits*. Science. 1995;(269):164-169.
3. Charlton BG. The scope and nature of epidemiology. *Journal of Clinical Epidemiology*. 1996;(49):623-626.
4. Silva LC. Lo individual y lo contextual en la búsqueda de factores etiológico. *Revista Humana*. 2002;6(3):117.
5. De Irala J, Martínez MA, Guillén F. ¿Qué es una variable de confusión? *Medicina Clínica*. 2001;(117): 377-385.
6. Silva LC. Estudios de casos y controles en psiquiatría: Causalidad, diseño y advertencias. *Actas Españolas de Psiquiatría*. 2004;32(4):236-248.
7. Silva LC, Barroso I. *Regresión logística*. Madrid: La Muralla; 2003.
8. Silva LC, Benavides A. Causalidad e inobservancia de la premisa de precedencia temporal en la investigación biomédica. *Revista Metodologica (Bélgica)*. 1999;(7):1-11.
9. Díaz OL. Factores de riesgo en la infección protésica vascular. *Revista Cubana de Higiene y Epidemiología*. 2000;38(1):29-36.
10. López S, et al. Factores de riesgo para la mortalidad de los pacientes quirúrgicos graves. *Revista Española de Anestesiología y Reanimación*. 2000;47(7):281-286.

11. Lleó A, et al. Factores clínicos de riesgo de la vitreorretinopatía proliferativa después de la cirugía del desprendimiento de retina. *Archivos de la Sociedad Española de Oftalmología*. 2000;(11):741-750.
12. Fuertes J et al. Factores de riesgo de asma, alergia e hiperreactividad bronquial en niños de 6 a 8 años. *Anales Españoles de Pediatría*. 2001;55:205-212.
13. Alegría A, et al. Evolución neurosensorial en recién nacidos de muy bajo peso de nacimiento a los 2 años de edad corregida. *Revista Chilena de Pediatría*. 2002;73(4):348-356.
14. Meza AM, et al. Factores de riesgo para el fracaso del tratamiento antituberculoso totalmente supervisado. *Revista de la Sociedad Peruana de Medicina Interna*. 2002;15(1):30-38.
15. Recabarren A, Cárdenas S. Factores de riesgo de asma infantil en niños que asisten al programa de control de asma del Hospital III Yanahuara ESSALUD - Arequipa. *Enfermedades del Tórax*. 2003;46(2):118-125.
16. Gimenoa D, Marko D, Martínez JM. Relación entre los factores de riesgo psicosociales laborales y la ausencia por razones de salud: motivos laborales y no laborales en España. *Archivos de Prevención de Riesgos Laborales*. 2003;6(3):139-145.
17. Baena JM, et al. Factores de riesgo modificables para el accidente isquémico transitorio no cardioembólico. Estudio de casos y controles en población general. *Revista de Neurología*. 2003;37(3):2-6.
18. Roselló M. Factores de riesgo asociados a glicemia elevada en ayunas en pacientes de la clínica de salud, de El Guarco de Cartago. *Revista Costarricense de Ciencias Médicas*. 2003;24(1-2):25-32.
19. Romero G, Martínez CA, Ábrego E, Ponce AL. Análisis multivariado de factores de riesgo asociados a la muerte fetal anteparto. *Calimed*. 2003;9(1):24.
20. Verdonck K, et al. Asociación entre infección por el virus linfotrópico humano de células T tipo I (HTLV-I) y mortalidad en pacientes hospitalizados con tuberculosis. *Revista Médica Herediana*. 2004;(15):197-202.
21. Rosas M, et al. Nueva visión de los factores de riesgo asociados a la hipertensión arterial en México. (Análisis por consolidación conjuntiva). *Archivos Cardiológicos de Mexico*. 2004;74 Supl 2: S164-178.
22. Llop JM, et al. Hipofosfatemia en nutrición parenteral: prevención y factores de riesgo asociados. *Nutrición Hospitalaria*. 2004;19(6):362-366.
23. Rosales M, et al. Incidencia y factores de riesgo de las infecciones quirúrgicas de by-pass vasculares en el hospital Juan Canalejo de A Coruña en el período 2000-2002. *Revista Española de Salud Pública*. 2005;(79):59-67.
24. Lizarralde E, et al. Pronóstico de las bacteriemias adquiridas en la comunidad ingresadas en un servicio de medicina interna. *Anales de Medicina Interna (Madrid)*. 2005;22(3):105-107.
25. Rodríguez A, et al. Relación entre el nivel socioeconómico y hábitos de vida, con el fibrinógeno y el factor von Willebrand en venezolanos sanos y con cardiopatía isquémica. *Investigación Clínica (Venezuela)*. 2005;46(2):157-168.
26. Ríos A, et al. Estudio multivariable de los factores de riesgo para desarrollar complicaciones en la cirugía del bocio multinodular. *Cirugía Española*. 2005;77(2):79-85.
27. Beaglehole R, Bonita R, Kjellström T. *Epidemiología básica*. Washington: Organización Panamericana de la Salud; 1994. (Publicación Científica N° 551)
28. Thompson B. Various editorial policies regarding statistical significance tests and effect sizes [en línea]. 2003 [fecha de acceso 20 de Marzo de 2003]. URL disponible en: <http://www.coe.tamu.edu/~bthompson>
29. Nickerson RS. Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*. 2000;5(2):241-301.
30. Comité Internacional de Directores de Revistas Médicas. Requisitos uniformes para la preparación de manuscritos. *Revista Panamericana de Salud Publica*. 2004;15(1):41-57.
31. Wilkinson L. Task Force on Statistical Inference. *Statistical methods in psychology journals*. *American Psychologist*. 1999;(54):594-604.
32. Sarria M, Silva LC. Las pruebas de significación estadística en tres revistas biomédicas: una revisión crítica. *Revista Panamericana de Salud Publica*. 2004;15(5):300-306.
33. Kleinbaum DG, Kupper LL, Muller KE, Nizam A. *Applied regression analysis and multivariable methods*. 3rd ed. USA: Duxbury Press; 1997.
34. Silva LC, Barroso I. Selección algorítmica de modelos en las aplicaciones biomédicas de la regresión múltiple. *Medicina Clínica*. 2001;(116):741-745.
35. Lord FM. On the statistical treatment of football numbers. *American Psychologist*. 1953;(8):750-751.
36. Guttman L. What is not what in statistics. *The statistician*. 1977;(26):81-107.

37. Leamer EE. Sensitivity analysis would help. *American Economic Review*. 1985;(75):308-313.
38. Rispau A, et al. Factores de riesgo asociados al consumo de antidepresivos. *Atención Primaria*. 1998;22(7):78-83.
39. Evans RG, Morris LB, Marmor TR. Why are some people healthy and others not? The determinants of health of populations. New York: Aldine Gruyter; 1994.
40. Breilh J. Nuevos conceptos y técnicas de investigación. Quito: Centro de Estudios y Asesoría en Salud; 1997. (Serie Epidemiología crítica; vol 3).
41. Silva LC. Cultura estadística e investigación científica en el campo de la salud: Una mirada crítica. Madrid: Díaz de Santos; 1997.
42. Park RL. Ciencia o vudu. De la ingenuidad al fraude científico. Barcelona: Gijalbo Mondadori; 2001.

Recibido el 28 de setiembre de 2005
Versión final presentada el 28 de octubre de 2005
Aprobado el 9 de noviembre de 2005