

Del *case report* a la evidencia epidemiológica de causalidad en investigación biomédica

Antonio Hernández Jerez

Arbor CLXXI, 675 (Marzo), 589-608 pp.

A lo largo de los últimos años la investigación clínica se ha desarrollado de manera significativa y ha contribuido a que actualmente las ciencias biosanitarias estén más fundamentadas en pruebas y evidencias científicas. La investigación en biomedicina se realiza a dos niveles, básico y epidemiológico. El enfoque epidemiológico presenta unos objetivos muy variados: desde la simple descripción de una realidad (caso clínico o case report) hasta la verificación de una hipótesis en relación con unos factores causales. En todo el espectro posible se pueden aplicar diferentes diseños epidemiológicos, para lograr la finalidad última que se persigue, aunque la validez de la elección determina el rigor científico de las conclusiones del estudio. Un elemento importante de cara a la inferencia de causalidad, sin posible participación de errores o sesgos, es la existencia de un grupo de comparación sobre el que el investigador no interviene (grupo no tratado o control) y, especialmente, la aleatorización del proceso, es decir, cuando la inclusión de los individuos en uno u otro grupo (tratado o control) depende exclusivamente del azar. El establecimiento de causalidad se realiza en base a las características cualitativas y cuantitativas de la asociación existente entre el proceso morboso y los factores de riesgo estudiados.

1. Introducción

La investigación nace de la curiosidad e inquietudes personales, observando hechos que no tienen explicación lógica o que son contradicto-

rios con las teorías clásicas. La investigación requiere dos condiciones básicas: observar críticamente la realidad para definir preguntas relevantes y aplicar un método lógico y reproducible para encontrar las respuestas a esas preguntas.

En el caso de la investigación biomédica, la interpretación de los resultados de los estudios clínicos y su posterior traslación a la práctica clínica está siendo actualmente objeto de debate. La profesión médica se está moviendo hacia lo que se denomina Medicina Basada en la Evidencia y la evidencia empleada en la toma de decisiones clínicas (tanto diagnósticas como terapéuticas) debe estar sustentada por métodos estadísticos fiables y fácilmente interpretables. Esta corriente científica, que ha inundado el quehacer clínico diario, se basa en que todo acto médico debe apoyarse sobre estudios científicos, fundamentalmente estudios de intervención, que hayan demostrado la efectividad de una determinada actitud terapéutica o prueba diagnóstica en el caso de que vaya a ser utilizada. De esta forma, la experiencia personal, la de otros compañeros o la opinión de expertos pasan a un segundo plano. No se trata de un movimiento nuevo, ni siquiera de una corriente revolucionaria, sino simplemente de una redefinición de los postulados científicos. De esta manera, todos los artículos publicados en la literatura científica son revisados metodológicamente y se establece su validez para que sus resultados y conclusiones puedan ser extrapolados a la población general. La parte fundamental de un estudio es el método utilizado para su realización ya que si el diseño no es el adecuado, todas las conclusiones que se deriven del mismo son erróneas y, por lo tanto, no aplicables a la población general.

Desde el punto de vista metodológico, la estrategia predominante para la identificación de causalidad descansa en la combinación inducción-verificación, planteamiento que nació como reacción al dogmatismo medieval. Se trata de un proceso racional que intenta encontrar explicaciones generales a partir de la observación de lo particular, tras lo cual se plantean hipótesis y, finalmente, se verifican. El enfoque inductivo establece, por una parte, que una teoría puede y debe emerger de la observación y, por otra, que dicha teoría ha de ser verificada contrastándola con la práctica de sus derivaciones. Avanzado el siglo XX, Popper realizó un planteamiento alternativo al inductivismo verificacionista y sostenía que la clave para avanzar no se halla en la verificación sino en su antítesis, la refutación. En rigor, Popper concede más interés a la hipótesis que más se aleje de lo que el conocimiento vigente permite esperar, ya que si no pudiera refutarse, tal resultado sería mucho más informativo. La lógica de este enfoque descansa en un hecho formalmente cierto: por

muchas confirmaciones que se consigan para un resultado, éste no queda por ello demostrado; un solo caso negativo, sin embargo, permitiría descartarlo como cierto.

Podría pensarse que el enfoque popperiano, al plantear que el progreso científico discurre a través de conjeturas y refutaciones y no mediante comprobaciones, no constituye realmente una visión opuesta al inductivismo. La negación de una hipótesis explicativa bien definida entraña una multitud de explicaciones alternativas. La confirmación de que una asociación es real (en el sentido de que no puede explicarse solamente por el azar) es, en efecto, equivalente a refutar la afirmación de que tal asociación es debida a la casualidad; pero no identifica a ninguna de las muchas otras explicaciones que pueda tener el hecho de haberla observado, tales como el efecto de uno o más sesgos, errores de medición, o la interferencia de factores de confusión. Popper y sus seguidores señalan, por tanto, que siempre pueden existir explicaciones alternativas y reclaman una permanente actitud de escepticismo que nos resguarde de la tentación de dar por cierto algo por el mero hecho de que se ha observado muchas veces.

Los inductivistas desdeñan el apego estricto a esta doctrina, considerada por ellos paralizante, a la vez que reivindican importantes éxitos en materia de causalidad. La historia de la ciencia biomédica les concede, en principio, una importante cuota de razón; por ejemplo, un éxito relativo del pensamiento inductivo lo constituyen las primeras hipótesis acerca de la causalidad infecciosa del SIDA. Si sólo fuera posible refutar hipótesis y no confirmarlas no existiría ninguna base sobre la que diseñar programas, elaborar normas, difundir mensajes sanitarios, emitir regulaciones o leyes. En cualquier caso, los investigadores adscritos a la corriente refutacionista, sean del ámbito clínico o epidemiológico, salen parcialmente del escollo, que supone la inevitable toma de decisiones médicas, diciendo que ésta ha de guiarse por la hipótesis que mejor resista los embates de las pruebas más exigentes.

Lo interesante es que, para la investigación biomédica y epidemiológica, tanto uno como otro enfoque reivindican la intervención de la estadística, a la vez que ninguno de los dos rechaza procedimiento estadístico alguno sobre la base de sus respectivos presupuestos teóricos. Por ejemplo, las pruebas o test estadísticos de contraste de hipótesis son vistas por unos como un instrumento corroborativo y por otros como un medio para la refutación. La epidemiología inductivista, desde luego, las aprecia desde la perspectiva ortodoxa: a partir de la muestra se sacan conclusiones sobre la población general y ese es el objetivo fundamental de la estadística inferencial.

2. Método científico

La ciencia avanza gracias a los esfuerzos investigadores que se realizan desde diversas áreas. Pero la investigación no puede ser caprichosa sino que debe seguir un orden racional que desde hace una serie de siglos viene denominándose método científico. Se entiende como tal al conjunto de procedimientos ordenados y sistematizados que se utilizan para adquirir conocimiento científico. Este método supuso un cambio muy importante en la forma de pensar y de entender la vida. Presenta dos propiedades importantes:

- 1.^a Objetivación: alcanzar una serie de datos y parámetros mediante el empleo de instrumentos objetivos de observación o medición. Hay que tener presente dos aspectos fundamentales: que el número de casos incluidos en el estudio sea suficiente para lograr dicha objetivación y, en segundo lugar, utilizar controles internos al experimento.
- 2.^a Generalización de los datos, es decir, la validez externa del experimento. Los resultados que se obtengan en el estudio deben ser aplicables no sólo a la muestra que se ha analizado sino a toda la población de referencia. En este sentido es necesaria la reproducibilidad de los datos, o sea, que otros investigadores lleguen a las mismas conclusiones en otros contextos y poblaciones diferentes, lo que podríamos llamar consistencia de los resultados.

Antes de que el método experimental irrumpiera en el mundo científico, el conocimiento humano se basaba en la *teología* (filosofía moral) y en la *metafísica* (filosofía especulativa). La base del conocimiento metafísico es la razón, lo importante es la lógica formal, la inferencia deductiva. Lo que legitima cualquier idea es el modo en que haya sido elaborada, esto es, un racionalismo cartesiano. Sin embargo, la lógica formal por sí sola no puede darnos a conocer la naturaleza. Para el científico no basta con que una hipótesis sea más o menos lógica, sino que hay que someterla al juicio de la experiencia mediante la aplicación del método. Si la hipótesis no supera la prueba ha de ser rechazada, sin importar que su mecanismo de elaboración sea lógicamente irreprochable. Esto nos conduce a la necesidad de recurrir a la experiencia para poder verificar la hipótesis. Nace de este modo la tercera forma de conocimiento humano, la *ciencia* o filosofía natural.

No existe una sola ciencia sino muchas, cada vez más, siendo el método aplicable a todas ellas. Estas ciencias se dividen en formales por acciones deductivas, es decir, aquellas que obtienen la verdad particular a partir de principios generales, y en fácticas por acciones inductivas, es

decir, la obtención de principios generales a partir de verdades particulares. Todo conocimiento científico tiene que alcanzarse a través del método. La medicina es una ciencia fáctica, de manera que todo aquello que no haya pasado por el método es pura hipótesis. De esta manera se diferencia de las ciencias formales (lógica y matemáticas), siendo éstas puramente deductivas que se ocupan de entes que no existen en la naturaleza sino en el mundo abstracto, producto del pensamiento humano. No obstante, algunos fenómenos naturales se pueden explicar en términos de leyes físicas o químicas que tienen una base matemática. Como la ciencia estudia las causas eficientes, rechaza las causas primeras, finales u ocultas, propias de la teología y de la metafísica.

El método científico tiene unas etapas claramente definidas y se caracteriza por ser sistemático, reflexivo, ordenado, crítico, universal e intemporal. Inicialmente, el investigador observa atentamente los hechos, a continuación formula una hipótesis que los explique conforme a los conocimientos científicos preexistentes, posteriormente trata de verificar la hipótesis mediante la experimentación o la observación estructurada y, por último, acepta, modifica o rechaza la hipótesis definida anteriormente. La etapa clave y diferencial del método epidemiológico es la verificación de la hipótesis mediante la experimentación, en la que el investigador realiza diferentes intervenciones a su voluntad, bien incrementando o reduciendo la intensidad de la hipotética causa para ver si la magnitud del efecto varía en la misma proporción, bien aplicando la causa contraria para ver si se produce el suceso contrario; o aplicando la misma causa a elementos diferentes. Las nuevas observaciones recogidas por medio de cualquiera de estos procedimientos permitiría verificar o refutar la hipótesis de trabajo.

Sin embargo, si la hipótesis que se quiere demostrar es de naturaleza etiológica, es decir, el investigador trata de analizar la relación causa-efecto entre un supuesto factor causal y una determinada enfermedad, el esquema experimental no puede realizarse en poblaciones humanas por razones éticas, prácticas y legales. Ha de recurrirse a la experimentación animal (cuyos resultados no son siempre extrapolables a la especie humana), o alternativamente a diseños observacionales que permitan analizar adecuadamente los fenómenos y circunstancias que acontecen de forma espontánea o habitual sobre humanos.

Gracias al método científico, a lo largo del siglo XX, se ha hecho imparable el avance de la Medicina, que inició su despegue exponencial a comienzos de la Edad Moderna, en el periodo postrenacentista. Así, por ejemplo, en el último medio siglo se han alcanzado más conocimientos que en los 2000 años anteriores. Una importante consecuencia de la apli-

cación del método científico es que ha permitido comunicar y divulgar los conocimientos generados, así como su posterior transferencia a la industria tecnológica permitiendo, de esta manera, el progreso de la humanidad y la mejora tanto de los procedimientos diagnósticos como terapéuticos. Ello ha redundado en un aumento tanto de la esperanza como de la calidad de vida de los hombres.

3. Tipos de estudios en investigación. Diseños epidemiológicos

La Epidemiología es la disciplina científica que comprende el estudio de la frecuencia, la distribución y los factores determinantes de las enfermedades en las poblaciones humanas. Se fundamenta en la utilización del método científico aplicado a los problemas de salud pública, es decir, en el denominado método epidemiológico.

Los diversos estudios epidemiológicos se basan sobre tres ideas implícitas:

1. Información sobre grupos. Aunque la práctica clínica diaria intenta verificar o rechazar diferentes hipótesis sobre la situación de salud de un paciente (lo que se denomina «diagnóstico»), el enfoque epidemiológico utiliza, sin embargo, información sobre la distribución de los problemas de salud y sus determinantes no en individuos sino en colectivos de personas, ya sean enfermas o sanas.
2. Comparación de grupos. Si bien en muchos casos es suficiente la descripción de la situación en un colectivo, generalmente se suelen realizar comparaciones con lo que ocurre en otros. Esas comparaciones son, a veces, implícitas, intuitivas, pero en la mayor parte de los diseños habrá que explicitar un grupo de comparación que además sea el más adecuado.
3. La distribución de la enfermedad. Los dos puntos anteriores se basan en una asunción previa: los problemas que afectan a la salud tienen una distribución no aleatoria, es decir, no se distribuyen al azar. Como consecuencia, es posible encontrar diferencias entre los distintos grupos que se comparan, diferencias que dependerán de la desigual distribución entre ellos de diferentes factores de exposición.

El abanico de posibles diseños es muy amplio y la elección de un determinado tipo de estudio estará condicionada, entre otras cosas, por la pregunta de investigación planteada y según los objetivos o hipótesis. Además, debe tenerse en cuenta el estado del conocimiento previo sobre el problema a investigar. Este será indudablemente un determinante del tipo de pregunta que se plantee y, por tanto, del tipo de estudio.

En líneas generales se puede decir que los dos grandes tipos de estudios epidemiológicos son los observacionales (o no experimentales) y los experimentales. Los *observacionales* son los más elementales pues no hay intervención del investigador en el diseño, ni aleatorización. Estos estudios pueden ser descriptivos (sobre individuos o sobre una población, pero no hay grupo de comparación) o analíticos (en este caso existe un grupo de comparación). Por el contrario, los estudios *experimentales* presentan la ventaja de que existe intervención en el diseño por parte del investigador y, además, aleatorización. Su utilidad es la verificación de hipótesis generadas en estudios analíticos previos.

Así, en una primera fase, nos encontramos con que lo importante es definir precisamente en qué consiste el problema, a pesar de que con frecuencia aparenta estar bien definido. Si el conocimiento existente sobre el problema identificado es escaso, el estudio más indicado será de tipo descriptivo. Se trata de «medir» el problema buscando diferencias entre los distintos grupos y subgrupos, en función de la desigual distribución de otras variables. En este caso, los estudios más adecuados son los ecológicos y transversales.

Cuando existen suficientes evidencias obtenidas a partir de estudios descriptivos previos se da un paso más y se intenta evaluar la posible asociación, de tipo causal, entre diferentes factores y el problema de salud de interés. En este momento de la investigación habrá que optar entre dos grandes grupos de estudios, los analíticos y los experimentales. En el primer caso están incluidos los estudios de seguimiento y los de casos y controles. La elección entre uno u otro dependerá de los objetivos, de criterios éticos, así como de los recursos y tiempo disponibles. En el grupo de los experimentales, se encuentran los estudios aleatorios controlados o ensayos clínicos y los diseños «cuasiexperimentales». El objetivo a esta altura de la investigación es la verificación de la hipótesis. Se intenta averiguar si hay una determinada asociación y si existen evidencias de que esa relación es de tipo causal, en cuyo caso podría hablarse de factores de riesgo.

4. Estudios no experimentales u observacionales

4.1. Diseños descriptivos

No cumplen ninguna de las tres condiciones del diseño experimental puro (hipótesis, grupo control y aleatorización), siendo el diseño más débil y por tanto es difícil extraer de él conclusiones válidas, ya que no se

pueden contrastar hipótesis. Algunos los consideran trabajos «no científicos» y otros, con mejor criterio, trabajos «cuasi científicos», ya que no se les puede negar todo el valor, siendo útiles en un gran número de circunstancias si están bien realizados. En líneas generales, describen el nivel de salud en una determinada población, estableciendo su distribución según diversas variables de persona, lugar o tiempo. Estos estudios permiten medir la importancia de un problema de salud, identificar posibles factores responsables conocidos e identificar grupos de riesgo. Entre estos estudios hay que mencionar los siguientes:

a) Caso clínico (*case report*). Estudio de un solo caso que por sus características excepcionales lo hacen merecedor de un alto interés científico. Es una forma muy popularizada de publicación en casi todas las revistas biomédicas. Obedece al interés de los profesionales de compartir su experiencia de diagnóstico y tratamiento con pacientes individuales. El hecho de que los casos clínicos no sean una muestra representativa de la población los priva de validez externa en términos de poder extrapolar los resultados a un universo. Sin embargo, estos estudios han contribuido de forma muy significativa al progreso de la medicina cuando se han realizado con el debido rigor metodológico.

b) Serie de casos. Recogen ordenadamente las observaciones comunes efectuadas en una serie de casos clínicos (agrupados alrededor de un mismo diagnóstico o problema clínico) recogidos por un mismo observador. Son muy habituales en las revistas médicas y, a menudo, ofrecen los primeros datos para caracterizar una enfermedad de reciente aparición (por ejemplo, el SIDA a principio de los años ochenta o más recientemente la variante humana de la enfermedad de las vacas locas). Algunos autores los denominan múltiples *case report* y son útiles si nos informan de hechos no conocidos hasta el momento en esa patología concreta, o si generan hipótesis etiológicas que puedan ser contrastadas en posteriores estudios. Se calcula que en España las publicaciones de este tipo suponen entre el 50 y el 80% de la producción científica médica, aunque su utilidad presenta una serie de limitaciones. En principio, no sirven para probar hipótesis estadísticas, dado que representan la experiencia de un solo observador o de un pequeño número de ellos. Desde un punto de vista epidemiológico presentan un problema crucial: la inexistencia de algún grupo que pudiera servir como comparación o control. Su validez en Medicina se basa en que contienen información de casos adquiridos a lo largo del tiempo y cuentan con pacientes pertenecientes a una serie que comparten algo en común (enfermedad, tratamiento, etc.). Incluso se pueden comparar de forma indirecta con series anteriores o con pacientes que hayan recibido distintos trata-

mientos para las mismas enfermedades, aunque con la lógica limitación de la falta de un grupo control.

c) Estudios ecológicos: estos estudios presentan una particularidad única consistente en que su unidad de análisis no son individuos en sí sino grupos de individuos. Requieren la existencia y disponibilidad de datos agregados, ambientales o globales, algo que no era disponible hasta hace poco tiempo, debido a la inexistencia de registros. La progresiva informatización de datos secundarios (ya existentes) está facilitando una enorme accesibilidad y, por lo tanto, la proliferación de estudios de tipo ecológico. La confidencialidad de las bases de datos, cada vez más demandada, impide en muchos casos el uso de información individualizada y obliga al uso de datos agregados. Estos diseños utilizan medidas que relacionan características de poblaciones en su conjunto y permiten describir los problemas de salud con respecto a elementos relacionados con características de las personas, el lugar y el tiempo. Así, se pueden cumplir objetivos tales como la rápida generación de hipótesis etiológicas o una fácil aproximación a la evaluación de programas preventivos o de intervención. Se clasifican en tres tipos: *a)* descriptivos o de mapas, en los que los datos se agregan por unidades geográfico-administrativas; por ejemplo, comparación de la prevalencia de SIDA por comarcas, provincias, comunidades autónomas, países, etc.; *b)* series temporales: describen el comportamiento de problemas sanitarios a lo largo del tiempo; por ejemplo, evolución de la mortalidad por cáncer a lo largo de los años; y *c)* correlación ecológica: analiza el grado de relación entre niveles promedio de exposición y el comportamiento del efecto hipotético, por ejemplo consumo promedio de tabaco y mortalidad por infarto de miocardio en distintas zonas o momentos del tiempo. En estos estudios puede presentarse un error de interpretación (denominado «falacia ecológica») si los fenómenos observados sobre una base colectiva se intentan extrapolar a individuos.

d) Estudios de prevalencia: son los más utilizados. Recogen información sobre la situación de salud de un grupo poblacional en un periodo concreto de tiempo establecido previamente en el diseño. La información se obtiene a través de encuestas específicas (encuestas de prevalencia) que, simultáneamente a los datos de interés, incorporan todo tipo de variables o exposiciones que se consideren oportunas. Estos estudios permiten cruzar información acerca de la prevalencia de problemas de salud con datos sobre la distribución de factores de exposición o de riesgo en un momento dado (corte transversal o «fotografía» de la situación). Si además se puede comparar esa descripción con lo que ocurre en distintos grupos, las diferencias que se encuentren servirán para generar hipótesis útiles en el diseño ulterior de estudios analíticos. Como se compara

simultáneamente exposición y presencia del efecto, surge el problema de si la exposición precedió a la enfermedad (en un esquema lógico de causalidad) o a la inversa. Otro problema radica en la muestra escogida, planteándose el dilema de si es representativa o no de la población general. Para que lo sea debe efectuarse un muestreo aleatorio de la misma, evitando incluir a voluntarios pues estos pueden alterar la representatividad de la muestra.

Como se puede ver, los estudios epidemiológicos descriptivos son estudios exploratorios, ya que sirven para generar hipótesis bien razonadas. Además, sus resultados pueden estimular la realización de otros estudios y servir de argumentación para desarrollar actividades de vigilancia y control biológicos en diferentes ámbitos sanitarios.

4.2. Trabajos analíticos

Son estudios de asociación ya que buscan demostrar que existe una asociación entre un determinado factor de riesgo y una o varias enfermedades, así como la magnitud de esa relación. Metodológicamente, se puede optar por seguir el desarrollo natural del proceso de enfermar y así clasificar a los individuos según estén o no expuestos al factor de riesgo en el que estamos interesados y, una vez definido un periodo de seguimiento, observar si la exposición influye sobre la aparición de la enfermedad; éste sería el fundamento de los estudios de cohortes. Otra aproximación metodológica consiste en clasificar los individuos según la presencia o ausencia de una enfermedad o alteración de la salud e investigar posteriormente si existió o no exposición al factor de estudio; este diseño se denomina estudio de casos y controles.

En el caso de los *estudios de seguimiento* el planteamiento es simple. Se parte de una o varias hipótesis que predicen la asociación entre una exposición y algún efecto sobre la salud. Si esa hipótesis es cierta, es razonable esperar que los individuos expuestos tengan una mayor probabilidad de desarrollar el efecto después de un periodo de tiempo más o menos largo. Estos resultados podrían compararse con los de otro grupo en el que no estuviera presente la exposición al factor de riesgo. Se dispondría entonces de, por lo menos, dos grupos: uno expuesto y el otro no; esos grupos se denominan cohortes. Ya solo queda seguir a ambas cohortes en el tiempo y verificar si efectivamente, en términos relativos, se produce un mayor número de nuevos casos en la cohorte expuesta en comparación con la no expuesta. En este diseño, a diferencia de los estudios experimentales, los investigadores no deciden qué individuos están expuestos y quienes no. Por distintas razones algunos individuos perte-

necen a alguna cohorte o a otra y sólo es posible observar qué ocurre a lo largo del tiempo. Como consecuencia, tampoco existe otra intervención que no sea la derivada de la observación y de la medición.

Los estudios de cohortes pueden ser de cohorte prospectiva (seguimiento) o retrospectiva (cohorte histórica), siendo estos últimos los más frecuentes en epidemiología laboral por ser más rápidos y menos costosos. Ambos tipos de diseño siguen una misma metodología, diferenciándose en el periodo de seguimiento. En los estudios de cohorte histórica es como si el investigador se trasladara en el tiempo y el periodo de seguimiento se iniciara mucho tiempo antes del inicio real del estudio. Los datos necesarios sobre la exposición y el nivel de salud que deben recogerse se obtienen de uno o varios registros, siendo por ello imprescindible disponer de registros documentales de buena calidad para poder realizarlos. Habitualmente son estudios de mortalidad, puesto que es muy difícil encontrar otros indicadores de morbilidad disponibles en registros.

En los estudios de *casos y controles* se parte de un grupo de individuos que presentan la enfermedad estudiada (casos) y otro libre de ella (controles) y se investiga retrospectivamente cuántos han estado expuestos al agente en estudio. Mediante un estudio de este tipo se identificó la asociación entre el consumo de cigarrillos y el desarrollo de cáncer de pulmón. Como medida cuantitativa de la asociación se calcula el exceso de riesgo de presentar la enfermedad en los expuestos al factor o agente nocivo con respecto a los no expuestos.

Los estudios de seguimiento plantean dificultades diversas, entre otras, el amplio tiempo de seguimiento que se requiere, especialmente cuando el análisis se centra en enfermedades con largos periodos de desarrollo (desde que se inicia la exposición hasta que aparece el efecto). Por ello los diseños del tipo casos-control son una opción frecuentemente preferida, ya que no presentan este problema.

Los diseños observaciones de tipo analítico son, pues, una buena opción cuando se intenta evaluar hipótesis, aunque no pueden aprovechar las ventajas de la asignación aleatoria y el control de las intervenciones, característico de los estudios experimentales. El problema va a ser, ya que no existe experimentación, en qué medida otros factores pudieran estar distorsionando la observación.

5. Estudios experimentales

Se engloban aquí un conjunto de diseños en los que existe una intervención planificada por los investigadores. El objetivo de un estudio

experimental es evaluar la eficacia de cualquier intervención, con fines terapéuticos, preventivos o rehabilitadores. La elaboración de un plan de investigación es uno de los aspectos más importantes del método experimental. Por esta razón, el diseño experimental debe ser escogido con sumo cuidado a la hora de plantear una investigación, ya que los errores que se cometan en esta fase difícilmente serán subsanables. Las condiciones que precisa un diseño experimental para ser considerado puro son, en sentido estricto, las siguientes: tener hipótesis de trabajo, disponer de un grupo control (de comparación) y asignación aleatoria de los participantes, es decir, garantizar que la distribución de la muestra entre el grupo control y el grupo de estudio se haya realizado al azar. Sin embargo, no todos los diseños cumplen estas tres condiciones.

La intervención consiste en la manipulación de una variable. Se somete un grupo a alguna exposición (por ejemplo un nuevo fármaco para tratar una enfermedad) y se compara con un grupo no expuesto que se usa como referencia o control (al que se le administra un placebo). La variable de exposición es la que se manipula y el objetivo es medir su efecto. Lógicamente la exposición consciente de seres humanos sólo es posible si se presume que tendrá efectos beneficiosos para ellos. Por eso este tipo de diseños tiene restringido su uso a evaluación de intervenciones preventivas, nuevos tratamientos o nuevas tecnologías diagnósticas. Los estudios experimentales son, por tanto, los únicos en los que la intervención está integrada en el diseño.

El ensayo clínico o *estudio aleatorio controlado* es el mejor diseño posible para verificar una hipótesis y el que cumple todas las condiciones necesarias para ser considerado un trabajo experimental; de hecho es lo más parecido a la experimentación pura en seres humanos, de ahí que sus conclusiones sean fácilmente interpretables y asumibles. Estas afirmaciones se sustentan en una serie de supuestos. En primer lugar, el experimento permite el control de la asignación. El hecho de incluir aleatoriamente a las personas en los grupos que se van a comparar (es decir, lo decide el azar, no el investigador) tiene una ventaja sobre el resto de estrategias consistente en que los grupos de comparación son muy similares entre sí (máxima comparabilidad). Es decir, para múltiples variables (como sexo, edad, etc.) es posible conseguir grupos homogéneos que presenten una distribución semejante de las variables distintas a la exposición que pudieran interferir en los resultados. De esta manera, se neutraliza el efecto de confusión que una distribución desigual podría introducir. Así, si se encuentran diferencias sólo quedaría una explicación: la intervención o manipulación de la variable.

Una opción intermedia es la que ofrecen los denominados *estudios cuasiexperimentales*. En ellos puede no existir asignación aleatoria individual sino que como mucho se decide al azar sobre qué grupo se va a intervenir (en estos estudios la composición de ambos grupos, control y experimental, nos puede venir dada). Una vez decidido esto, solo las personas que integran este grupo recibirán la intervención. Incluso es posible el diseño cuasiexperimental en el que no exista grupo control, sino que se evalúa la situación en un solo grupo antes y después de la intervención. Se utilizan en epidemiología para estudiar efectos de un agente nocivo y, dada la finalidad del trabajo, los grupos control y experimental están definidos previamente, de manera que los individuos no son asignados aleatoriamente a un grupo o a otro. Por el contrario, en un estudio experimental, los individuos se distribuyen al azar en dos grupos y sobre uno de ellos se efectuará la intervención novedosa, que interesa contrastar, mientras que en el otro se realizará la intervención estándar preexistente.

Aún cuando los estudios experimentales pueden parecer enormemente atractivos por sus ventajas, en multitud de ocasiones son impracticables. Por un lado, razones éticas impiden exponer a seres humanos a situaciones que pudieran ser perjudiciales para ellos (por ejemplo, ensayo de un nuevo medicamento sobre mujeres embarazadas). Por otro lado, también presenta limitaciones científicas: la aleatorización no garantiza la homogeneidad de ambos grupos, efecto que sólo está presente cuando la muestra sea de tamaño suficientemente grande. Por tanto, la homogeneidad no podrá darse por supuesta si el número de factores que pueden influir en el resultado es grande y la muestra no es muy numerosa, no pudiéndose aplicar la ley de los grandes números. Por último, cada vez es más frecuente que la investigación etiológica se dirija hacia exposiciones ligadas a formas y estilos de vida; en este caso, cualquier modificación potencialmente beneficiosa puede ser imposible de conseguir porque los individuos asignados aleatoriamente no estarían dispuestos a modificar sus hábitos y costumbres (por ejemplo, dejar de fumar, cambiar de dieta, hacer deporte).

De todo lo expuesto se desprende que en clínica humana es difícil hacer diseños experimentales puros, ya que sus limitaciones son al menos tan grandes como sus ventajas. Suelen quedar reservados para estudiar problemas de gran trascendencia, en los que se pretende evaluar mejoras cuantitativamente moderadas pero científicamente importantes. En estos casos está justificada la gran inversión en número de pacientes, dinero y tiempo, asumiendo sus limitaciones como precio a pagar por investigar con seres humanos.

6. Población y muestra

Una vez organizado el diseño experimental, el investigador debe establecer los criterios de selección de los sujetos que van a formar parte de su estudio. Por tanto, no trabajará con toda la población sino que escogerá un número determinado de casos a los que se denomina muestra. Los resultados obtenidos con esa muestra serán los que posteriormente se generalicen extrapolándolos a toda la población mediante cálculos estadísticos apropiados.

Se define «muestreo» o extracción de muestras al proceso de seleccionar parte de una población para que la represente a toda ella. Una muestra será por tanto un subconjunto de unidades que comprenden la población. Los científicos generalmente trabajan con muestras y no con poblaciones, ya que es más económico y eficaz hacerlo así, con grupos pequeños de elementos y no con todo el conjunto. Por ello, las muestras son formas prácticas y eficaces de obtener datos.

Se define población como toda la agregación o cúmulo de casos que cumplen con un conjunto predeterminado de criterios. Se distinguen dos tipos de poblaciones: población diana o blanco, que es aquella correspondiente al conjunto de individuos al que hace referencia la pregunta a efectuar y que es centro del estudio; sería la población a la que se desearía generalizar los resultados. En segundo lugar, está la población de estudio accesible, constituida por el conjunto o agregado de casos que cumplen con los criterios predeterminados y que es accesible al investigador como el conjunto global de sujetos a estudiar.

Para conseguir una muestra homogénea, el investigador utiliza la definición de población y los criterios de inclusión y exclusión. El criterio de inclusión más importante es la definición de la enfermedad que se estudia. Dicha enfermedad se puede definir sobre la base de diversos parámetros, debiendo el investigador elegir los más aceptados al respecto. Para ello resulta de gran utilidad la bibliografía. Es importante utilizar, siempre que sea posible, más de un criterio para definir la enfermedad. Los criterios deberán quedar establecidos de un modo claro, objetivo y preciso para evitar posteriores dudas. Con respecto a los criterios de exclusión, algunos pacientes que hayan sido definidos previamente deberán ser excluidos del estudio sobre la base de dos consideraciones fundamentales: por presentar condiciones que supongan un riesgo adicional para ese paciente o por presentar algunas características que puedan alterar los resultados que buscamos. Por ello, se deberán eliminar del estudio los pacientes cuyas circunstancias específicas puedan modificar la respuesta o fenómeno que vamos a analizar.

Una de las propiedades de los seres vivos es la variabilidad, lo que los convierte en una población heterogénea respecto a muchos factores (sociales, económicos, profesionales, físicos, culturales, etc.). Sin embargo, el investigador está interesado en que la población de estudio sea razonablemente homogénea para que de esta forma sus resultados se vean afectados sólo por el factor a evaluar. El encontrar una variabilidad excesiva en un estudio nos podría explicar por sí sola las diferencias encontradas en el mismo. No obstante, esto es un arma de doble filo. Si se establecen unas condiciones basales muy estrictas (utilizando los criterios de inclusión y exclusión correspondientes), la población resultante será muy homogénea, pero difícilmente representará a la población general, heterogénea por definición. Los resultados obtenidos serán fácilmente interpretables, pero habremos limitado seriamente el alcance del estudio, pudiéndose aplicar las conclusiones sólo a los escasos individuos que cumplan los requisitos impuestos en las condiciones basales. Caso de establecer unos criterios excesivamente laxos, nos encontraremos con el problema contrario.

Por ello, debemos buscar un equilibrio entre homogeneidad y la representatividad. Los criterios de selección no deberán ser muy laxos ni estrictos, sino exactos y suficientes. Los criterios de inclusión no deberán ser muy restrictivos, ni los de exclusión demasiado numerosos. Debemos apoyarnos en la bibliografía para solventar dentro de lo posible este problema. Es importante mantener esta norma general: en caso de duda se debe dar prioridad a la validez interna del estudio (homogeneidad) sobre la validez externa (representatividad). Con ello deberemos excluir a un gran número de pacientes de nuestro estudio, siendo la población resultante menos representativa que la que vemos en la práctica diaria, pero este es un precio a pagar por investigar con seres humanos, variables y heterogéneos por definición.

En cuanto al *tamaño de muestra*, todo investigador debe estudiar un número suficiente de casos para obtener resultados clínicamente relevantes, pero no tanto para malgastar tiempo y dinero, así como causar molestias innecesarias a los pacientes. Cuando se utilizan técnicas probabilísticas es posible conocer de antemano el tamaño necesario de la muestra para lograr un nivel específico de precisión. Sin embargo, no es posible señalar una sola fórmula debido a que el tamaño de la muestra a estudiar depende de varios factores: el grado de precisión requerido, lo heterogéneo de los atributos en investigación, las técnicas de muestreo que utilicemos y el tipo de técnica de estimación usado. Vuelve a ser necesaria en este apartado la bibliografía para orientarnos. Si la bibliografía no nos saca de dudas habrá que calcular estadísticamente qué ta-

maño debe tener nuestra muestra en función de la variabilidad del parámetro que queramos estudiar, así como de la precisión de la estimación y el nivel de confianza deseado. En otras ocasiones habrá que recurrir al llamado «estudio piloto» o preliminar, realizado a base de una pequeña muestra.

La investigación en medicina clínica necesita de estudios experimentales muy potentes para detectar posibles avances. Caso de no disponer de un número de pacientes suficientes y requerir de una muestra grande, se puede recurrir a diferentes procedimientos, entre ellos los denominados estudios multicéntricos, en los que colaboran varias instituciones separadas geográficamente entre sí. De esta forma se obtendrán resultados que quedan fuera del alcance de un trabajo individual aislado. Sin embargo, estos estudios presentan dos inconvenientes: la pérdida de la unidad de juicio por participar diferentes investigadores y la disminución de la homogeneidad de la muestra y el aumento de forma paralela de la dispersión de los datos.

7. Asociación y causalidad

Las investigaciones clínicas se suelen clasificar en etiológicas y evaluativas. Las primeras se ocupan de estudiar las relaciones causa-efecto, mientras que las evaluativas valoran nuevos métodos diagnósticos o terapéuticos disponibles. Los estudios o investigaciones etiológicas intentan analizar e interpretar los diversos índices de aparición de enfermedades, estableciendo una relación causal entre el factor de estudio y las posibles modificaciones que se observen en el criterio de evaluación. Esto es, índice de aparición entre individuos expuestos y controles que no están expuestos al factor etiológico sospechoso. Este tipo de trabajos necesita de un grupo control, pues en caso de no poseerlo nunca se podría establecer una relación causa-efecto, todo lo más una asociación. Este es el principal motivo por el que un diseño descriptivo no es aplicable a este tipo de investigaciones.

Una relación entre dos factores puede ser falsa, casual o causal. Las falsas asociaciones suelen estar motivadas por diseños defectuosos que conducen a conclusiones erróneas y que pueden evitarse con una correcta planificación. La asociación casual o por azar puede aparecer cuando tanto la enfermedad como la exposición tienen relación con un tercer factor, detectado al medir la exposición al supuesto agente etiológico. Cualquier investigador, que pretenda establecer una relación causa-efecto, debe pensar con anterioridad si existe alguna otra explicación para los

hechos que han ocurrido que no sea la directa de causa y efecto. Si se descarta la relación casual, suponemos que la asociación es causal.

La asociación causal de un estudio para contrastar una hipótesis necesita demostrar no sólo que un determinado factor contribuye a la aparición de una enfermedad, sino que su eliminación disminuye la incidencia o gravedad de la misma. Los métodos estadísticos ayudan a este tipo de tarea pero no se bastan por sí solos para garantizar una relación causal (podría si el estudio fuera aleatorizado, pero no es posible éticamente).

7.1. *Criterios adicionales en que se basa la relación causa-efecto*

El estudio de relación causa-efecto se basa fundamentalmente en descartar explicaciones alternativas, como podría ser el papel del azar o la contribución de factores de confusión o errores de método. Cuando faltan estas explicaciones alternativas, la interpretación de una relación causa-efecto se basa en unos criterios que se enumeran a continuación:

- **Fuerza de asociación.** Cuanto mayor sea la medida de asociación que utilicemos, mayor será la probabilidad de que el factor sea causa del efecto. Cuanto más intensa sea, más verosimilitud adquiere la hipótesis subyacente. Ello se debe a que una de las incertidumbres estriba en que la correlación observada pudiera ser sólo un reflejo del efecto que ejercen otros condicionantes; si la asociación es intensa, la probabilidad de que pueda explicarse enteramente por el efecto de tales factores (los llamados factores de confusión) es mucho menor.
- **Consistencia de la asociación.** El hecho de que diferentes investigadores en poblaciones distintas hayan obtenido resultados similares da fuerza a la idea de causalidad. Lo contrario no significa que esta asociación no sea causal, sino que habría que tener en cuenta la metodología empleada y la calidad de cada estudio.
- **Relación temporal entre el factor de estudio y el efecto.** Significa que la causa debe preceder al efecto. Aunque este principio sea evidente, en los estudios transversales y en algunos diseños de casos y controles esta condición no siempre se puede asegurar.
- **Existencia de un gradiente biológico.** La observación de una sostenida relación dosis-respuesta (a mayor dosis del factor, mayor intensidad del efecto o de la respuesta) aumenta el grado de confianza que se deposite en la validez del silogismo causal.
- **Plausibilidad y coherencia biológica de la asociación.** La hipótesis debe ser verosímil. La esperanza de que la plausibilidad biológica no sea exigida es la levadura de muchas expresiones de la pseudo-

ciencia. Por otro lado, los resultados obtenidos tienen que ser compatibles (coherentes) con los conocimientos anteriores sobre la historia natural y biológica de la enfermedad. Esta regla es discutible, ya que en realidad todo resultado científico es provisional y los nuevos paradigmas que mencionara Kuhn (1975) se forjan sobre la base de la irreverencia hacia la obra de los antepasados.

- **Indicios experimentales.** En la experimentación pura el investigador modifica la realidad según un plan y registra lo que ocurre como resultado de la evaluación del efecto de su intervención. En los últimos años se ha reavivado el debate entorno a la utilidad de los estudios no experimentales para demostrar relaciones de causalidad. El pensamiento epidemiológico predominante defiende que los esfuerzos de investigación que no hacen uso de la experimentación son, a pesar de ello, de extrema utilidad ya que contribuyen a fundamentar empíricamente las hipótesis. Últimamente se advierte una desenfrenada producción de trabajos observacionales (sobre todo estudios casos y control) destinados a la identificación de factores de riesgo. Además de que tales esfuerzos nunca darán respuestas definitivas, esta práctica parece estar generando más confusión que esclarecimiento.
- **Analogía.** Una relación causa-efecto sale fortalecida cuando existe una relación similar en otra situación. Si otros factores siguen caminos causales análogos a los de la hipótesis que se valora, la convicción de que sea válida se incrementa. Este criterio puede ser útil en algunas circunstancias, pero dista de constituir una demanda, a diferencia de lo que ocurre con la experimentación.

8. Precisión y validez de los estudios epidemiológicos

Aunque sea el diseño ideal, no siempre es posible realizar estudios experimentales, con grupos de tamaño suficiente. La mayoría de los estudios epidemiológicos se basan en la observación de determinados fenómenos en uno o varios grupos de individuos. El tamaño del grupo estudiado, por una parte, y las limitaciones derivadas de la mera observación de un grupo de individuos determinado, por otra, hacen que los resultados que se obtienen en los estudios epidemiológicos no estén exentos de limitaciones. Estas limitaciones se refieren habitualmente a dos aspectos importantes: la precisión y la validez.

La *precisión* está en relación con el tamaño del grupo o muestra, de manera que el grado de precisión de los resultados obtenidos es tanto

más elevado cuanto mayor sea el tamaño del grupo estudiado. Dado que los problemas de precisión están ligados al tamaño de la muestra, la importancia de los mismos y sus consecuencias pueden conocerse con antelación.

Los problemas de *validez*, por su parte, pueden ser más sutiles y difíciles de prever, por lo que representan un mayor reto para la adecuada interpretación de los resultados. En Epidemiología se dice que un resultado es válido cuando está libre de errores sistemáticos. El error sistemático o sesgo, a diferencia del error aleatorio, no depende del tamaño muestral, es decir, no se atenúa aumentando la muestra estudiada. El error sistemático puede introducirse en diferentes fases del estudio y tiene como consecuencia la producción de resultados sesgados. Estos problemas de validez pueden surgir: *a)* al incluir en el estudio muestras de personas que no son representativas de la población de referencia que realmente se desea estudiar, al menos en aquellas variables que son objeto del estudio (sesgo de selección); *b)* por errores en la información sobre la exposición, la enfermedad u otras condiciones en los grupos que se comparan (sesgo de información), y *c)* por falta de comparabilidad de los grupos en relación con variables que, si bien no son objetivo central del estudio, están asociadas con ellos y pueden confundir los resultados (sesgo de confusión). El epidemiólogo está obligado a evitar o controlar los errores que pueda haber cometido en su estudio y, en todo caso, a evaluar su impacto cuando interpreta sus resultados. En resumen, antes de concluir sobre la validez de los resultados de un estudio hay que descartar que éstos se deban al azar, a sesgos de selección, a sesgos de información o al efecto confundente de terceros factores.

Las referencias a la validez realizadas hasta ahora se refieren a la validez de los resultados del estudio para la población objeto del mismo, lo que se denomina también validez interna. Por el contrario, la validez externa se refiere al hecho de poder extrapolar los resultados, es decir, si los resultados obtenidos se pueden aplicar o generalizar también a otras poblaciones similares. Lo fundamental es que la validez interna es un requisito indispensable para la validez externa.

Bibliografía

- BURGOS RODRÍGUEZ, R. (1998): *Metodología de investigación y escritura científica en clínica*. Granada: Escuela Andaluza de Salud Pública.
- GARCÍA CRIADO, E. I. (2001): Investigación en atención primaria. *Semergén* 27, 190-211.

- KUHN, T. S. (1975): *La estructura de las revoluciones científicas*. México, D.F.: Fondos de Cultura Económica.
- PIÉDROLA, G., REY CALERO, J., DOMÍNGUEZ, M., CORTINA, P., GÁLVEZ, R., SIERRA, A., SÁENZ, M. C., DOMEZ, L. I., FERNÁNDEZ-CREHUET, J., SALLERAS, L., CUETO, A. y GESTAL, J. J. (1994): *Medicina Preventiva y Salud Pública* (9.ª ed.). Barcelona: Ediciones Científicas y Técnicas S.A.
- PRAT, A. y SALLERAS, L. (1995): El método epidemiológico en salud laboral. En *Manual de salud laboral*, P. Sanz-Gallén, J. Izquierdo y A. Prat Marín (Eds). Barcelona: Springer-Verlag Ibérica.
- ROTHMAN, K. (1987): *Epidemiología moderna*. Madrid: Díaz de Santos.
- RUIZ FRUTOS, C., ASÚA BATARRITA, J., y FERNÁNDEZ AJURIA, A. (1997). Epidemiología laboral. En *Salud laboral. Conceptos, técnicas para la prevención de riesgos laborales*, F. G. Benavides, C. Ruiz Frutos y A. M. García García (Eds). Barcelona: Masson.
- SHAKESPEARE, T. P., GEBSKI, V. J., VENESS, M. J. y SIMES, J. (2001): Improving interpretation of clinical studies by use of confidence levels, clinical significance curves, and risk-benefit contours. *Lancet* 357, 1349-1353
- SILVA, L. C. (1997): *Cultura estadística e investigación científica en el campo de la Salud. Una mirada crítica*. Sevilla: Díaz de Santos.