

Investigación en atención primaria

E.I. García Criado

Médico de Familia. Centro de Salud del Carpio. Zona Básica de Bujalance. Córdoba. Vocal Nacional de Investigación de SEMERGEN.

EN QUÉ CONSISTE EL MÉTODO EXPERIMENTAL

El enfoque científico es el método más avanzado de adquirir conocimiento que el ser humano jamás haya usado. En él se combinan características inductivas y deductivas, a fin de crear un sistema que aporte conocimientos más fiables que los que aportaría la tradición, la autoridad o la simple experiencia.

Se conoce por método de investigación o método experimental al conjunto de procedimientos ordenados y sistematizados que se utilizan para adquirir conocimiento científico. Está constituido por una serie de normas o reglas que en gran parte son difíciles de entender y a veces mal conocidas. Se precisa, por tanto, para su correcto manejo experiencia, algo de imaginación y una gran habilidad. En rasgos generales, el método se basa en observar los hechos, elaborar hipótesis que se puedan acomodar a ellos y plantear experiencias que los confirmen o refuten, observando si tienen lugar las consecuencias previstas.

El método experimental avanza de acuerdo con el esquema de la figura 1.

1. Objetividad y verificabilidad. El investigador para obtener conocimientos científicos utiliza una serie de datos y parámetros conseguidos a través de los sentidos (medidas, pesos, contajes, etc.), se basa pues en la percepción sensorial. Todos podemos ser víctimas del engaño que esta percepción provoca, por lo que debemos esforzarnos en todo momento para que los datos que utilizamos al investigar sean lo más imparciales y objetivos posibles.

La ciencia cuenta con medios para aumentar la objetividad de un testimonio sensorial individual, entre los cuales están:

- Utilizar medidas los más exactas posibles, con patrones aceptados internacionalmente (registro de cantidades y no de cualidades).
- Uso de instrumentos precisos y apropiados.
- Utilizar controles (si es posible, internos al experimento).
- Que el número de casos estudiado sea suficiente.

– Usar criterios objetivos para juzgar los hechos, y que éstos sean lo más independientes posible del juicio del observador.

La reproducibilidad o verificabilidad posibilita que cualquier persona que actuara en igual forma y condiciones que nosotros obtenga los mismos resultados.

2. Generalización de los datos. El investigador debe generalizar los datos obtenidos, para que sus conclusiones sean útiles y aplicables a los demás investigadores. En medicina clínica esto se consigue gracias a la estadística, que nos marca el grado de probabilidad o grado de credibilidad de las informaciones que queramos transmitir. De esta forma la ciencia rechaza la hipótesis que tienen una probabilidad pequeña de ser ciertas, aceptando por el contrario las que tienen un alto grado de credibilidad.

¿ES NECESARIO EL MÉTODO DE INVESTIGACIÓN EN ATENCIÓN PRIMARIA DE SALUD?

A lo largo de los siglos el médico siempre ha tratado de extraer de su quehacer diario las consecuencias prácticas oportunas, destinadas al mejor cuidado de sus pacientes. No se puede dudar la validez del conocimiento adquirido a través de este proceso antiguo, natural y venerable, que siempre ha estado ligado a nuestra profesión médica. Sin embargo, la experiencia que el médico va acumulando a lo largo de su vida no es fácil de comunicar ni transferir, por lo que la medicina progresó poco como ciencia. Esto hace que aunque sea muy útil para el desempeño individual de la profesión, no lo sea tanto en el ámbito colectivo.

Por otra parte, la historia de la medicina nos enseña que, en muchas ocasiones, la experiencia individual e incontrolada puede conducir a errores muy graves, dada la benevolencia humana para juzgar sus propias opiniones y resultados. Sirva como ejemplo Paracelso y la iatroquímica que probablemente mató más personas que curó.

En la evolución de la historia debemos esperar al siglo XVI para que Francis Bacon anuncie el método experimental, aunque en realidad fuera Galileo, en el siglo XVII, y aplicado a la astronomía, el que lo desarrolle por primera vez. Posteriormente se aplica el método en física mecánica y óptica (Newton), fisiología (Harvey), química (Lavoisier), biología (Linneo), etc., no sin algunas críticas por parte de sus coetáneos.

Correspondencia: Dr. E.I. García Criado.
Avda. del Aeropuerto, 35, 4.-3.º. 14004 Córdoba.
Correo electrónico: egarcia@semegen.es

SEMERGEN: 2001; 27: 190-211.

Observamos, pues, como diferentes disciplinas se fueron haciendo científicas. Pero debemos esperar al siglo XIX para que la medicina clínica aplique el método científico en los hospitales de París (Bichat, Laennec, Corvisart, etc.), basándose en la idea anatomo-clínica de Morgagni, del concepto localizado de la enfermedad. Se precisaba de una obra en la cual la vieja "anatomía práctica" del *Sepulchretum* y la narración de historias clínicas con el correspondiente protocolo de autopsia, fuese elaborada con amplitud y rigor suficientes. Esto se consiguió con el magnífico libro *De sedibus et causis morborum per anatomen indigatis*, publicado por Giovanni Battista Morgagni. Con ello, no sólo se describe la lesión, sino que se establecen "géneros lesionales", esto es, tumor, rotura, humores, tensiones etc. Esto conlleva a un avance de la medicina desarrollando elementos de medida como termómetro, esfigmomanómetro, fonendoscopio, otoscopio, laringoscopio, etc.

El siglo XX presenta como novedad el desarrollo de la terapéutica, haciendo imparable el avance de la medicina. Sirvamos como ejemplo práctico el duplicar la esperanza la vida media de la población (gracias a la aparición de la medicina preventiva, terapéutica, etc.), o bien el hecho de que en medio siglo se desarrollaran más conocimientos que en los aproximadamente 2.300 años precedentes desde Hipócrates. Ello es debido a la aplicación del método de investigación, y a que éste presenta una forma transferible y comunicable de obtener conocimientos, aunque siempre haya estado ligado a la medicina hospitalaria. Frente a la aplicación del método en el ámbito hospitalario, la asistencia primaria, fortalecida tras la conferencia de Alma-Ata (1978), inicia sus propios caminos y campos de investigación, sobre todo epidemiológicos, aspectos organizativos de la asistencia, sistemas de registro, etc. Esto nos hace pensar que no es el hospital el único lugar donde se pueda investigar y que, por tanto, "existe vida inteligente fuera de los hospitales".

Otro motivo para conocer el método es la necesidad de mantener nuestros conocimientos al día, a fin de mejorar la calidad asistencial. Para ello es preciso una lectura crítica de los trabajos publicados, que nos informarán de los sesgos y limitaciones que cada uno presenta. Podríamos preguntarnos: ¿qué artículo es más útil para la toma de decisiones clínicas?; ¿qué defecto presenta un trabajo?; caso de tenerlo, ¿pueden afectar o limitar seriamente las conclusiones?; estas conclusiones ¿son aplicables a nuestros pacientes?; etc.

A modo de resumen, coincido con la idea de Argimón al decir que se debe insistir en la urgente necesidad que tiene la atención primaria de salud de contar con proyectos de investigación numerosos y de calidad, que den lugar no sólo a publicaciones más o menos brillantes en revistas de prestigio, sino también a progresos reales en campos que le son propios y que la diferencian como área científica de la asistencia hospitalaria. Sería preciso trabajar con vistas a mantener investigaciones con posibilidad de futuro a medio y largo plazo, y evitar el artículo fácil que aumente nuestro currículo y que ayuda muy poco al saber médico.

Debido a que generalmente el método de investigación no se suele impartir durante la enseñanza de pregrado y posgrado, es mi intención desarrollar, mediante este tema central de la revista, el estudio del método y su aplicación en la atención primaria de salud, labor que espero tenga buena acogida por parte de los lectores.

¿QUÉ ES LA CIENCIA?

Como se ha comentado anteriormente, el método experimental, también llamado método científico o método de investigación, supuso un cambio mental muy importante en la historia de la humanidad.

Hasta el momento en que hizo su aparición, el conocimiento humano se basaba en la teología (filosofía moral) y en la metafísica (filosofía especulativa). La metafísica es considerada una parte de la filosofía que trata del ser en cuanto tal, y de sus propiedades, principios y causas primeras. El nombre de metafísica fue dado por los comentaristas de Aristóteles al tratado filosófico que éste llamó filosofía.

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 con 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, 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 comprobar las hipótesis. Nace de este modo la tercera forma de conocimiento humano, la *ciencia o filosofía natural*.

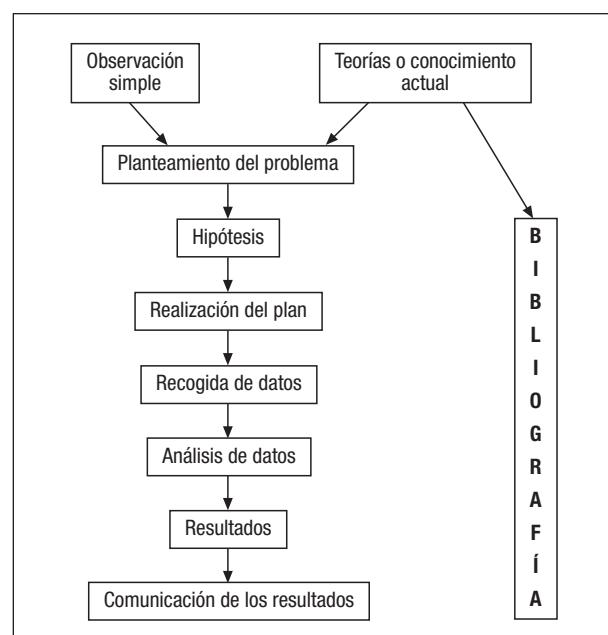


Figura 1. Esquema de las etapas de una investigación.

Podemos definir la Ciencia como "el conocimiento cierto de las cosas por sus principios y causas, o bien el cuerpo de doctrina metódicamente formado y ordenado, que constituye un ramo particular del humano saber".

Todo conocimiento científico ha de ser hallado a través del método, siendo apoyado éste por las ciencias fácticas (recordemos que fáctico proviene de *factum*, que significa relativo a los hechos, en oposición a lo teórico). La medicina es una ciencia fáctica, de modo que todo lo que en medicina no haya pasado por el método es pura hipótesis. Dicha hipótesis podría tener una buena carta de presentación, dependiendo de quién lo comunique, pero no dejará de ser una opinión y no formará ciencia. Se diferencia de esta manera de las ciencias formales (lógica y matemática), siendo éstas puramente deductivas y ocupándose de entes que no existen en la naturaleza, esto es, entes abstractos como números, integrales, derivadas, etc.

Al estudiar la ciencia las causas eficientes, rechaza las causas primeras, finales u ocultas, propias como dijimos de la teología y de la metafísica.

Nos encontramos ante una ciencia que no cree en el principio de autoridad, aplicándolo sólo de forma atenuada y en contadas ocasiones (opiniones de autoridades científicas). Dicha ciencia coincide en esto con la filosofía moderna (desde Descartes o desde Kant, con su filosofía crítica), al no aplicar el principio de autoridad. Sin embargo, no lo hace con la escolástica medieval, que lo aplicaba, y aun menos con la teología, que continúa haciendo uso de ella en la actualidad (sirva como ejemplo la autoridad de quien revela: Dios). De aquí que la base del conocimiento teológico sea la fe unida en mayor o menor grado a la razón.

Por último, es necesario hacer notar que las ciencias aumentan su número de forma vertiginosa, llegando en la actualidad a superar las 2000, y siendo el método aplicable a casi todas ellas (a excepción de algunas como amor y religión). Estas ciencias se dividirán 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.

PASADO, PRESENTE Y FUTURO DE LA INVESTIGACIÓN EN ATENCIÓN PRIMARIA

Como comentamos, la investigación nace de la curiosidad e inquietudes personales, observando hechos que no tienen una explicación lógica o bien que son contradictorios a las teorías clásicas. Según Juan Gérvás la investigación requiere dos condiciones básicas: *a*) observar críticamente la realidad para definir preguntas interesantes, y *b*) aplicar un método lógico y reproducible para encontrar las respuestas a esas preguntas.

El *pasado* de la investigación en atención primaria en nuestro país mostraba un pobre bagaje. Sirva como ejemplo que entre 1973 y 1984 de los 47.202 artículos publicados sólo el 0,59% estaban relacionados con este sector. Para José Antonio Martínez este déficit podría ser atribui-

ble a la gran presión asistencial que el médico de primaria soportaba, así como a la falta de apoyo económico y estímulos por parte de la Administración sanitaria, que tenía como objetivo primordial la asistencia sanitaria.

En la *actualidad* queremos pensar que la situación está en vías de cambio progresivo, pasando el peso porcentual de publicaciones referentes a atención primaria a ser del 4,1% respecto al total. De igual modo ha aumentado el número de proyectos del Fondo de Investigaciones Sanitarias (FIS) que se conceden en los últimos años. No obstante la investigación sigue ocupando un porcentaje muy pequeño entre las prioridades de la Administración sanitaria que continúa volcada con la parcela asistencial.

El *futuro* de la investigación se prevé halagüeño en tanto y en cuanto está demostrado que una buena base investigadora aporta a los servicios de salud calidad, efectividad y eficiencia en la atención. Mediante dicha investigación se pueden modelar las bases de la atención primaria a fin de desarrollar métodos que midan el estado de salud, la causalidad y gravedad, el cumplimiento del paciente, los resultados de la atención a una población, la calidad de dicha atención o la necesidad de derivación. Este tipo de investigación, denominada por Starfield como básica, sería la piedra angular del futuro de una mejora en los servicios asistenciales a nuestros pacientes.

¿A QUIÉN INTERESA LA INVESTIGACIÓN EN LA ATENCIÓN PRIMARIA DE SALUD?

La investigación en atención primaria interesa a los actores que de algún modo desempeñan un papel relevante dentro del funcionamiento de la sanidad y éstos son:

1. *Los pacientes*: les supone cambios en el resultado en términos de salud o de satisfacción.
2. *Los gestores y planificadores de servicios de salud*: enfatiza la importancia en que la toma de decisiones en el acto médico tenga una base científica que la apoye.
3. *Profesionales asistenciales*: por la satisfacción de investigar y el poder participar en las decisiones de cómo se ha de hacer una asistencia basada en lo que nos ha demostrado ser efectivo y eficiente y no sólo en lo que sea más barato.

¿CÓMO PODRÍAMOS MEJORAR LA INVESTIGACIÓN EN ATENCIÓN PRIMARIA?

En palabras de Gómez de la Cámara estamos ante una "paradoja asistencial", fenómeno que se observa en atención primaria, donde "se dispone de un montón de evidencias científicas que no consiguen dotar al sistema de la efectividad deseada". Ante un gran número de ensayos clínicos, hay un enorme grado de ineffectividad en el sistema, por ejemplo el pobrísimo grado de control de lípidos, hipertensión o diabetes, o bien de resistencias bacterianas que tenemos en nuestro medio. La causa primordial de esta situación sería la inadaptación al medio y al perfil de los individuos que estamos atendiendo en nuestras consultas.

Esto se podría solventar si utilizáramos el ensayo clínico (en fase 3 y 4), como instrumento de investigación y generación de evidencias científicas. No obstante podremos chocar de frente con las prioridades del Sistema Nacional de Salud que se sitúan en la minimización de costes, relación coste-beneficio y el coste-utilidad en los productos utilizados en atención primaria como base fundamental para su actuación y líneas maestras.

Desde 1997, se realizaron en España un total de 533 ensayos clínicos de los cuales sólo 23 correspondieron a atención primaria.

Las últimas tendencias del FIS y siguiendo las directrices del vigente Plan de Investigación, Desarrollo e Innovación Tecnológica 2000-2003, son:

1. Investigación básica orientada.
2. Investigación en medicina clínica, fisiopatología y terapéutica.
3. Investigación en epidemiología y servicios de salud.
4. Investigación sociosanitaria.

A estos campos tenemos que unir un creciente interés en el área de telediagnóstico, de carácter novedoso en todos los ámbitos de la asistencia sanitaria.

Basándonos en la revisión de Ricoy et al (1996) de los proyectos financiados por el FIS en el período 1988-1995, de las tres áreas analizadas, las ciencias básicas, la investigación clínica y el área de la epidemiología, salud pública y servicios de salud, es esta última la que tiene menos proyectos financiados y, sobre todo, la que recibe menos financiación.

Dentro del volumen de proyectos recibidos provenientes de atención primaria es de menos del 2% del total de proyectos y del importe en el período que analiza. Sin embargo, el porcentaje de proyectos del área de epidemiología, salud pública y servicios de salud que proceden de atención primaria ascienden al 14%. Esto nos viene a mostrar el alto porcentaje de proyectos que se financian en atención primaria del área de servicios de salud, tendencia que se ve reflejada en las constantes publicaciones y comunicaciones en dicho ámbito de estudios descriptivos.

No obstante, no debemos sentir menoscabo por que el gran aporte a la ciencia investigadora desde la atención primaria sean los estudios descriptivos, toda vez que de ellos se pueden obtener los datos necesarios para crear futuras hipótesis de trabajo que contrastaríamos realizando otros tipos de trabajos de investigación. Sírvanos como ejemplo de la importancia de este tipo de estudios que una de las revistas internacionales de mayor prestigio como *The Lancet* publica un 56% de estudios descriptivos, dándonos con ello a entender que no se permitiría el lujo científico de incluir en sus páginas tan alto porcentaje de artículos de mala calidad.

En los últimos años vemos cómo van aumentando el número de estudios analíticos en atención primaria si bien y según opinión de Juncosa Font podríamos hacer más atractiva la investigación mediante la búsqueda de temas más próximos a la atención primaria, que respeten los va-

lores intrínsecos de nuestra filosofía, evitando la deslealtad hacia la profesión. Dichos temas demostrarían su utilidad y sobre todo nos facilitarían la accesibilidad a dicha investigación con una clarificación de horizontes que pasen por mejorar la recogida de los datos, implicar a otras disciplinas afines y como último requerimiento y por ello no menos importante, incentivando profesional y económicamente al médico investigador. Esto, junto con facilitar una aplicación más directa de los resultados obtenidos en nuestras investigaciones, sería más que suficiente para potenciar al máximo la capacidad investigadora del médico de atención primaria.

¿QUÉ TIPO DE INVESTIGACIÓN HACEMOS EN ESPAÑA Y QUÉ DEBEMOS INVESTIGAR EN ATENCIÓN PRIMARIA?

Los temas de investigación en nuestro país, según una revisión de Marset Campos de los publicados en los últimos 23 años serían:

1. Trabajos clínicos: 63,5%.
2. Organización y administración: 14%.
3. Promoción de la salud y prevención de enfermedades: 13,5%.
4. Salud pública: 7,5%.
5. Metodología y docencia: 6,7%.

Los estudios contrastados en los últimos años van más dirigidos a aspectos ligados a control de calidad y a administración y programación, situación que podría venir dada por las líneas de la administración derivadas de la reestructuración económica del Estado de bienestar que primar estos estudios sobre los de promoción y prevención de la salud.

Respecto a la producción científica por comunidades autónomas (1984-1994), son las de mayor demografía como Cataluña, Madrid, Andalucía y la Comunidad Valenciana, junto con las más extensas, como Castilla y León, Aragón, País Vasco y Castilla-La Mancha, las que más producen. No obstante si valoramos los indicadores de trabajos por cada 100 médicos de atención primaria, Cataluña, seguida de Aragón, son las que más producen. Quedaría en un tercer lugar Madrid seguida de Navarra y Murcia. Si esta catalogación la hiciésemos para el conjunto de la medicina convertida en trabajos por 100 médicos, arrojaría una imagen distinta, siendo Navarra la primera región, seguida de Madrid y Cataluña.

Ante los datos anteriores nos surge la pregunta de ¿qué debemos investigar en atención primaria? La investigación debería basarse en los mismos principios de investigación que la de otros ámbitos, con la salvedad de especial contenido de nuestra asistencia.

Ya en 1962 el Colegio de Médicos Generales de Inglaterra publicó la "Guía para la investigación en medicina general", en cuyo prólogo se podía leer: "sois vosotros, los médicos generales, quienes debéis ser los pioneros de la investigación médica. Estáis en condición de observar el

comienzo de una enfermedad, seguirla a través de sus etapas, y verla en su verdadera perspectiva".

No obstante, para poner límites lógicos a la recomendación en las líneas de investigación en atención primaria, y siguiendo las acertadas directrices del FIS, sería recomendable que fueran dirigidas a los siguientes campos:

1. *Investigación básica.* Investigación aplicada, descripción o análisis de la estructura de proceso o resultados. Inclusión del desarrollo de instrumentos para medir la calidad de vida de los pacientes, la satisfacción de nuestros usuarios, etc.

2. *Investigación epidemiológica.* La realización de grandes estudios epidemiológicos se vería dificultada por la escasez en las bases de datos poblacionales de datos sociosanitarios. Ello nos llevaría a limitar a localidades o áreas pequeñas nuestro ámbito investigador; no obstante, la privilegiada situación dentro de la esfera biopsicosocial del paciente nos va a permitir estudiar problemas de alta prevalencia y el estudio natural de la enfermedad, mejorando con ello la efectividad de nuestra actividad al dirigir los recursos a aquellos que más lo necesiten.

3. *Investigación clínica.* Nuestra especial situación nos permite mejorar la validez de muchos estudios, mediante la potenciación de la validez externa (grado en que la población de estudio es similar a la población general). Del igual modo, los ensayos clínicos con diseños muy controlados serán los idóneos para establecer la eficacia de un fármaco o la puesta a punto de cualquier tipo de tratamiento en las diversas fases de la investigación.

4. *Investigación en servicios sanitarios.* Orientaremos nuestros esfuerzos hacia el estudio del nivel de efectividad en la presentación de cuidados a nuestra población, como pueden ser los modelos de funcionamiento de los equipos, calidad de atención, evaluación de contenidos y periodicidad de los programas, modelos de gestión, evaluación económica, etc.

Visto lo anterior disponemos en atención primaria de un amplio campo para investigar, y parafraseando al sher-

pa Tensing cuando se le preguntó cómo él y Sir Edmund Hillary llegaron al Everest, diríamos que la investigación se consigue: "De modo muy sencillo: simplemente poniendo un pie delante del otro y manteniendo clara la meta".

EL MÉTODO EXPERIMENTAL APlicado A LAS CIENCIAS BIOMÉDICAS

El objeto del estudio de la medicina es el ser humano y por esta razón la investigación biomédica tiene necesidad de ciertas consideraciones éticas y legales. La primera de estas consideraciones es cuestionarnos la ética de la investigación con seres humanos. La respuesta indudable es afirmativa siempre que tengamos en cuenta lo siguiente:

1. Cada tratamiento médico es un experimento no controlado, y no cuesta nada controlarlo de acuerdo con los principios del método experimental.

2. Si elegimos no experimentar avanzaremos por medio del sistema de prueba y error, con lo cual retardaremos el avance de la ciencia y correremos riesgos no calculados.

Por tanto, debemos considerar ética la experimentación con seres humanos siempre que esté bien diseñada y ejecutada, y siempre que no atentemos contra la dignidad del hombre ni contra su libertad de decisión sobre su vida y bienestar (tabla 1). Estos aspectos quedaron plasmados en la Declaración elaborada en Helsinki (1964) por la XVIII Asamblea de la Asociación Médica Mundial, modificada y perfeccionada posteriormente en Tokio (1975) y en Venecia (1983). Su antecedente inmediato fue el llamado Código de Nuremberg, desarrollado en 1947 (tabla 2).

La Declaración de Helsinki esta aceptada por todos los editores de revistas biomédicas del mundo y supervisada por los Comités de Investigación existentes en la mayoría de los hospitales, se trata de normas de obligado cumplimiento cuya trasgresión nos transporta fuera de las fronteras de la ciencia. La función de los comités de investigación ética clínica está referida en la tabla 3.

Tabla 1. Características éticas en los ensayos clínicos en atención primaria. Cuestiones éticas en medicina de familia

1. Sólo se aceptará utilizar un placebo como control cuando no exista intervención o terapia mejor que ésta
2. Debe existir una tercera persona (diferente al médico y al paciente) con capacidad para monitorizar el estudio y paralizarlo si fuera necesario
3. Se debe utilizar ensayos multicéntricos para disminuir la duración del tratamiento y el número de sujetos precisos
4. Se podría solicitar si es preciso la participación en el estudio de otro individuo diferente al médico de familia
5. Cuando informemos al paciente que se incluye en el estudio debemos explicarle que:
 - El futuro de la atención médica no se verá afectado por su decisión de participar en el estudio
 - La explicación para la obtención del consentimiento debería ser realizada por otro individuo que no sea el médico de familia que atiende al paciente evitando con ello el compromiso previo
 - El paciente nunca se le deberá pagar por incluirlo en el estudio y caso de hacerlo nunca debería de ser el médico el que lo haga
6. Las cuestiones éticas en medicina de familia son:
 - Tipo I. Cuestiones de actitud profesional
 - Visión de la atención primaria
 - Alcance del compromiso con el paciente
 - Tipo II. Cuestiones de decisión (dilemas)
 - Problemas de decisión entre opciones
 - Tipo III. Cuestiones de ejecución
 - Dificultades en la puesta en práctica por obstáculos internos o externos

Tomada de Coulehan JL, Kenneth J, Block M. Ethics of clinical trials in family medicine. J Fam Practice 1985; 21: 217-222, y modificada por el autor.

DISEÑO EXPERIMENTAL

Consideraciones generales

El objetivo de un estudio experimental es evaluar la eficacia de cualquier intervención con fines preventivos, rehabilitadores, curativos, etc.

La elaboración de un plan de investigación es, por tanto, 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 plantearnos una investigación, ya que los errores que cometamos en esta fase difícilmente serán subsanables.

A continuación, vamos a tratar de conceptos básicos del diseño experimental, intentando con ello hacerlo grato y accesible al investigador así como orientarlo para evitar los tan temidos “errores de diseño”.

Las condiciones que precisa un diseño experimental para ser considerado puro son, en sentido estricto, las siguientes:

- Tener hipótesis de trabajo.
- Tener un grupo control.
- Garantizar que la distribución de la muestra entre el grupo control y el grupo de estudio se haya realizado al azar, esto es, de forma aleatoria. Sin embargo no todos los diseños cumplen estas tres condiciones. Dependiendo de ello nos encontraremos con diferentes tipos de diseños experimentales.

Clasificación de los diseños clínicos

Desglosaremos los trabajos clínicos en:

1. Trabajos 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 difícil de extraer de él conclusiones válidas, ya que no se pueden contrastar hipótesis. Para algunos autores son trabajos “no científicos”. Otros autores, quizás con mejor criterio, les denominan trabajos “cuasi científicos”, ya que no se les puede negar todo el valor, siendo útiles en un gran número de circunstancias, si están bien hechos. Dentro de este grupo de trabajos destacaremos:

a) Caso clínico (*case report*). Estudio de un solo caso clínico que por sus características excepcionales le hacen merecedor de un alto interés científico (sírvanos como ejemplo un intento de innovación diagnóstica o terapéutica en una situación clínica desesperada).

Tabla 2. Código de Nuremberg

1. Consentimiento voluntario
2. Diseño adecuado y justificado
3. No causará daños innecesarios y nunca se practicará si provoca lesión o muerte
4. Su riesgo no será mayor que la importancia del problema en estudio
5. Dirigido por personal científicamente competente
6. El sujeto es libre para poner fin al experimento
7. Debe finalizarse si hay sospecha de que cause daño

b) Series de casos. 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). Algunos autores los denominan “estudios de cohorte” o 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 el futuro.

Su validez en medicina se basa en:

– Contener información adquirida a lo largo del tiempo (estudios longitudinales y no transversales).

– Contar con pacientes pertenecientes a una serie que comparten algo en común (enfermedad, tratamiento, etc.).

– Poderse comparar de forma indirecta con series anteriores, o con pacientes que hayan recibido distintos tratamientos para las mismas enfermedades, aunque con la lógica limitación de la falta de un grupo control.

2. Trabajos analíticos. Son trabajos que contienen una hipótesis expresa, que precisa para su contrastación algún tipo de control. Al no llevar grupo control, no pueden ser considerados experimentales. Debido a ello se les denomina “trabajos de investigación no experimentales” o “trabajos descriptivos usados analíticamente”. Para otros son los trabajos “cuasi científicos”.

Aunque son estudios no controlados, por definición, la mayoría llevan incorporado un “control histórico”, esto es, la posibilidad de comparar los resultados de nuestra serie con otras publicadas en la bibliografía, si bien el control histórico es débil y presenta una serie de inconvenientes como son:

– Comparar resultados de dos o más centros diferentes, que pueden estar alejados geográficamente y con poblaciones no equiparables, respecto a cultura, información sanitaria, nivel de salud, etc.

– Comparar con nuestros propios resultados anteriores, sin tener en cuenta los cambios en la historia natural de la enfermedad, criterios de clasificación más evolucionados, o mejoras en los cuidados médicos. Son estas causas suficientes para explicar las diferencias, caso de encontrarlas.

Sin embargo, los trabajos analíticos son útiles y están justificados para informar del resultado espectacular de

Tabla 3. Funciones de los comités éticos de investigación

1. Garantizar la calidad científica y la relevancia de la investigación
2. Garantizar la competencia de los investigadores
3. Garantizar la protección de los participantes en ensayos clínicos:
 - a) Minimizar los riesgos
 - b) Asegurar la selección equitativa de los sujetos
 - c) Procurar que el consentimiento informado sea debidamente obtenido y documentado
 - d) Controlar la seguridad de los sujetos
 - e) Respetar la confidencialidad
 - f) Comprobar la existencia de un seguro de cobertura por los daños que se puedan producir a los sujetos

una nueva terapéutica (caso de la penicilina en el tratamiento de la meningitis meningocócica), o bien para reportar observaciones de toxicidad o cualquier otro problema importante con una terapia de uso habitual. Sus ventajas radican en que son más baratos y necesitan menor número de pacientes que los controlados, son fáciles y rápidos de realizar y no crear problemas éticos a los investigadores.

Una forma útil de minimizar los inconvenientes del control histórico sería:

1. Utilizar controles recientes (evitamos que el paso del tiempo y con él los avances médicos influyan en nuestro estudio).
2. Utilizar, en lo posible, controles propios (pacientes tratados en el mismo hospital con criterios de selección, zona geográfica y patrones de referencia parecidos).
3. Asegurarse de que ambas series sean razonablemente comparables (edad, sexo, clínica, etc.).

Para la interpretación de los resultados de este tipo de estudios debemos tener en cuenta las diferencias. Si estas son muy importantes, los resultados son totalmente válidos (límites de confianza del 90%). Si, por el contrario, no lo son, debemos ser cautos a la hora de interpretarlas, utilizándolas como nuevas hipótesis que serían contrastadas por futuros estudios controlados. Por último hay que ser muy cuidadoso al intentar establecer por este mecanismo una relación causa-efecto, ya que este tipo de estudio suele requerir un grupo control.

Estudios controlados: retrospectivos y prospectivos

Además de contar con grupo control, los diseños experimentales se pueden realizar de forma prospectiva o retrospectiva, existiendo también otros diseños controlados como son los apareados, cruzados, estratificados, etc. A continuación se desarrollan los tipos de diseños experimentales antes mencionados:

Estudios retrospectivos controlados

En este tipo de estudio, los dos grupos se establecen sobre la base de las historias clínicas de los pacientes. Estos trabajos son baratos, rápidos y carecen de problemas éticos. No obstante este diseño presenta graves inconvenientes, entre los que destacamos:

1. La información que contiene la historia clínica no se obtuvo con el propósito de investigar, lo que conlleva a la falta de criterios uniformes y de estandarización a la hora recoger dicha información.
2. La muestra no viene dada, siendo su representatividad dudosa. Nos encontramos, por ello, con dificultad al definir los criterios por los que dicha muestra fue seleccionada.
3. Carecen de aleatoriedad en la distribución de los pacientes entre los grupos control y experimental.
4. No se puede garantizar la fiabilidad de los registros, ante la imposibilidad de realizar un control de calidad retrospectivo.

Debido a esto sus conclusiones son menos consistentes que las obtenidas mediante un diseño prospectivo.

A pesar de estos inconvenientes, pueden estar justificados en las siguientes circunstancias:

1. Estudios de enfermedades raras, en las que el conseguir una muestra con un tamaño adecuado de forma prospectiva nos llevaría mucho tiempo.
2. Para evaluar diferencias a largo plazo, entre dos terapias en uso (si las diferencias que pretendemos encontrar son a corto plazo, el uso de un estudio retrospectivo sería, cuando menos, sospechoso, salvo que la enfermedad en cuestión sea muy rara).

Utilidad de los trabajos y valoración de las diferencias de los estudios retrospectivos controlados

La utilidad fundamental de los estudios retrospectivos radica en la carencia de problemas éticos. A dichos estudios, a diferencia de los realizados mediante control histórico, se les pueden aplicar test estadísticos para establecer si las diferencias encontradas son significativas. Si bien contamos con el problema de que sólo podemos dar valor definitivo a las grandes diferencias. Las pequeñas diferencias generarían nuevas hipótesis, contrastables con estudios prospectivos posteriores. Todo ello hace que se utilicen mucho en investigación clínica y formen buena parte de los trabajos actualmente publicados en revistas.

Precauciones a tomar en los estudios retrospectivos controlados

Las limitaciones inherentes al diseño nos obligan a tomar una serie de precauciones para garantizar un resultado razonablemente adecuado:

1. Debemos estar seguros de que las historias clínicas recogen con fiabilidad los datos a estudiar, utilizando sólo datos fuertes, poco subjetivos y muy importantes para la evolución del paciente, que no hayan podido pasar desapercibidos fácilmente al elaborar la historia clínica.
2. Las historias clínicas deben ser completas al respecto que se estudia, para ello se analizarán varias historias al azar, rechazando las que presenten datos incompletos. Si el número de historias rechazadas supera el 10%, el estudio probablemente debe ser abandonado.
3. Si se compara una nueva terapéutica con otra en uso, utilizar como control la terapia más ampliamente aceptada, no cualquier otra.
4. El investigador debe intentar que ambos grupos (el control y el experimental) sean tan homogéneos como sea posible, para ello deberá establecer unas condiciones básicas estrictas y similares para ambos.

Estudios prospectivos controlados

Es el auténtico trabajo experimental. Por medio de este diseño el investigador contrasta su hipótesis mediante la comparación de los resultados obtenidos en el grupo control y en el grupo experimental.

Dentro de este tipo de estudios y dependiendo de cómo se formen ambos grupos podemos hablar de:

1. Estudio de intervención. Sería aquel en el que la composición de ambos grupos (control y experimental), nos puede venir dada. Algunos autores lo consideran un diseño “cuasiexperimental”.

Se utilizan en epidemiología para estudiar efectos de un agente nocivo, siendo la propia finalidad del trabajo la que impide la asignación aleatoria al grupo control y al grupo experimental.

2. Estudio prospectivo, controlado y aleatorizado. Es el diseño más perfecto y el único que cumple todas las condiciones necesarias para ser considerado un trabajo experimental. Lo más parecido a la experimentación pura en seres humanos y en palabras de Carrasco de la Peña “estamos combatiendo al azar con el azar, y ya más no se puede hacer”.

Se denomina también ensayo clínico aleatorio (*controlled clinical trial*) y sus conclusiones son fácilmente interpretables y asumibles. Se reserva para la evaluación de diferentes terapias, aunque puede ser aplicable a ciertos estudios operacionales (política administrativa, gestión sanitaria, etc.).

La asignación de cada paciente a uno de los dos grupos del estudio se suele realizar mediante “randomización”, término (cuyo nombre deriva de su promotor, Random) que indica sorteo aleatorio, esto implica que ni el investigador ni el paciente pueden influir en la composición de los grupos. Es, por tanto una tabla de número aleatorios (obtenida de un libro de estadística), o una lista de dichos números elaborada por un ordenador la que realiza la distribución. Con ello tenemos las máximas garantías de que los grupos son homogéneos y, por tanto, comparables.

Problemas e inconvenientes que plantea la aleatorización

Aunque estos estudios son considerados por muchos científicos como el *gold standard* de la investigación clínica, plantean algunos problemas e inconvenientes que analizaremos a continuación:

1. Limitaciones éticas importantes:

– Justificación del trabajo. Para que un investigador adopte este diseño es necesario que desconozca por completo cuál de las dos terapias es mejor. Si estudios controlados previos han establecido alguna diferencia sustancial entre las terapias, ya no esté éticamente justificada la aleatorización, pues enviaríamos a un grupo de pacientes a la peor terapia.

– En general, se acepta que hay que informar al paciente de que su terapia va a ser elegida a sorteo y éste debe aceptarlo (Declaración de Helsinki). De no aceptarlo se excluirá del estudio. No se considera, por tanto, ético el cambio de grupo o la exclusión una vez realizado el sorteo.

– Existe un problema importante y de difícil solución cuando una terapia se muestra claramente superior a la contraria, pero la diferencia no alcanza aún significación estadística. El dilema está servido; si detenemos el estudio hemos perdido tiempo y dinero, junto con la imposibilidad de obtener una conclusión científica válida. Si por el

contrario no lo hacemos y continuamos el trabajo, estamos ante una situación ética incómoda para el investigador. Existe un término medio que sería mantener al investigador clínico en la ignorancia de los resultados parciales, aunque ello no sea siempre posible.

2. Limitaciones científicas:

– La aleatorización nos garantiza la homogeneidad de ambos grupos, pero este efecto sólo está presente cuando la muestra tiene un tamaño suficiente. De este modo, 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. Ante este supuesto hay que volver a evaluar la distribución de los factores en ambos grupos y evitaremos sorpresas desagradables de última hora (consideraremos una variable desigualmente distribuida cuando la tasa de variación supera el 25% esperable por azar).

– Debemos tener en cuenta el entrecruzamiento de pacientes entre ambos grupos médicos. Distinguiremos dos tipos de entrecruzamientos; aquellos pacientes asignados a un grupo de tratamiento médico que en el curso del estudio necesiten del tratamiento quirúrgico por agravamiento o complicaciones de su patología, y un segundo tipo de entrecruzamiento por la preferencia expresa del paciente. No obstante estos supuestos no alteran altamente el resultado del estudio si su proporción no alcanza cifras significativas.

3. Otros inconvenientes. Dentro de este punto englobaremos:

– Consumen mucho tiempo y esfuerzo.

– Son muy costosos económicoamente.

– Exigen un gran número de pacientes por perderse algunos en los seguimientos (se pretenden demostrar diferencias muy pequeñas). Se considera que las pérdidas para el seguimiento deben ser inferiores al 5% de la muestra total, no debiendo nunca superar el 10%.

Sin embargo el problema real no está en el número de pérdidas, sino en las causas que las motivaron. Por ejemplo, la pérdida de un paciente por un cambio de médico, residencia, etc., no influiría en las conclusiones del trabajo. Por el contrario, si la pérdida es debida a efectos secundarios de la medicación o complicaciones graves, puede producirse un sesgo en favor o en contra de la hipótesis del trabajo.

Por todo lo expuesto, vemos 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 pretendemos evaluar mejorías 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.

Otros diseños controlados

En los diseños controlados que hemos estudiado hasta el momento el investigador compara los resultados obtenidos en el grupo control y en el grupo experimental como si se tratara de dos muestras independientes, y a esto es lo que llamamos *diseño paralelo*.

Hay situaciones en las que por trabajar con muestras pequeñas, nos puede interesar controlar alguna de las variables a estudiar para evitar una distorsión de los resultados, evitando una desigualdad en la distribución de ambos grupos de estudio. Cuando nos encontramos ante esta situación estaremos utilizando un *diseño no paralelo*. Dentro de este tipo de diseño diferenciaremos:

1. **Diseño apareado.** El apareamiento (*matching*) consiste en seleccionar, para cada individuo del grupo de estudio, uno o varios sujetos de referencia similares en las características que deseamos controlar. De forma aleatoria, uno integrará el grupo control y otro el grupo experimental. Si apareamos por un gran número de variables corremos el riesgo de no disponer del suficiente número de sujetos adecuados, al tiempo que puede enmascararse la asociación de interés, por tener grupos excesivamente parecidos. Esto es lo que se denomina sobreapareamiento u *over-matching*.

2. **Diseño cruzado.** El estudio cruzado o secuencial es un diseño muy útil para evaluar nuevos fármacos. En este tipo de diseño no es necesaria la asunción de que el análisis se realizará una sola vez para calcular los niveles de significación, permitiéndonos las normas para finalizar el estudio una evaluación continua de los datos.

Se establecen dos grupos de estudio no independientes. El grupo A comienza con tratamiento X y luego cambiará al Y. El grupo B recibirá también ambos tratamientos pero en orden inverso. La homogeneidad de la muestra no nos plantea ningún problema toda vez que cada paciente actúa como su propio control (trabajando con muestras menores que los diseños paralelos).

Presenta limitaciones ya que sólo es aplicable a enfermedades crónicas, y ello siempre que contemos con el denominado "tiempo de blanqueo", esto es, el tiempo suficiente para evitar que persistan los efectos del primer tratamiento y que estos puedan solaparse con los del segundo, alterando, por tanto, los resultados.

3. **Diseño estratificado.** Se trata de dividir la muestra en subgrupos o estratos mutuamente excluyentes y lo más homogéneos posible respecto a las variables de confusión. Dentro del estrato los sujetos son asignados por azar al grupo control y al experimental.

La estratificación nos permite formar subgrupos similares con respecto a un factor pronóstico determinado y hacer las comparaciones adecuadas, aumentando la potencia estadística del estudio.

Como inconveniente encontramos que al aumentar el número de subgrupos y de pacientes de forma geométrica, nos podemos quedar con pocos pacientes en cada estrato, lo que dificultaría el análisis estadístico. Para evitarlo debemos estratificar con respecto a covariables real-

mente importantes, esto es, hacer los estratos que sean necesarios, pero los mínimos posibles.

Estudios epidemiológicos

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 son aquellos/as que intentan analizar e interpretar los diversos índices de aparición en 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 de la enfermedad entre individuos expuestos y testigos o controles que no están expuestos al factor etiológico sospechoso.

Es lógico pensar que este tipo de trabajos necesita de un grupo control, caso de no poseerlo nunca se podría establecer una relación causa-efecto, cuanto más sólo una asociación. Este es el principal motivo por el que un diseño descriptivo no es aplicable a este tipo de investigaciones.

Sólo en contadas ocasiones y en circunstancias muy especiales podremos hacer uso de un control histórico en una investigación etiológica. Como ejemplo tenemos la determinación de accidentes mortales antes y después de implantar la obligatoriedad del uso de cinturón de seguridad en los automóviles. La legislación nos impide utilizar testigos o controles en este caso. Cuando nos encontramos ante estas circunstancias aceptaremos las limitaciones propias de los estudios con control histórico referidas anteriormente.

Estudios retrospectivos controlados

Dentro de los diseños experimentales podemos utilizar estudios retrospectivos controlados, en los que los pacientes con la enfermedad serían los casos y los que no la presenten los controles. Por *odds* se entiende la razón entre la probabilidad de que un suceso ocurra y la probabilidad de que no ocurra. Si la probabilidad de que una persona con la enfermedad esté expuesta es de 0,60, la *odds* de exposición se calculará dividiendo este valor por la probabilidad de que no esté expuesta ($0,60/[1-0,60] = 1,5$). Se considera *odds ratio* a la razón entre la *odds* de exposición observada en el grupo de casos y la *odds* de exposición en el grupo control. Esto es, la medida de asociación o de riesgo de padecer un determinado problema de salud asociado a la presencia de una exposición. Dicha *odds ratio* nos permite, por tanto, valorar la importancia de la causa.

Aunque los estudios retrospectivos aporten resultados en un corto espacio de tiempo y permitan evaluar de forma simultánea varios factores de riesgo, presentan múltiples inconvenientes. Entre los más importantes cabe citar:

1. Dificultades de selección de un grupo adecuado.
2. Sesgos de memoria del entrevistado en cuanto a la exposición al factor etiológico sospechoso.
3. Sesgos propios del observador al recoger la información pertinente.

Por ello debemos reservar los estudios retrospectivos para evaluar enfermedades poco frecuentes con causa sospechada relativamente frecuente, o bien procesos patológicos con largos períodos de latencia (p. ej., la investigación de la relación existente entre el consumo de tabaco y la incidencia de carcinoma de pulmón).

Estudios prospectivos

Se denominan también estudios de cohorte o estudios longitudinales. Al ser la aleatorización en estos casos inaceptable (por tratarse de estudios etiológicos con agentes nocivos), los estudios longitudinales se basan en reunir una cohorte o grupo estudio que se encuentre libre de la enfermedad considerada, y vigilarla hacia delante para observar la aparición de dicha enfermedad en alguno de sus miembros. Como ejemplo clásico de este tipo de estudio tenemos el tan conocido estudio Framingham cuyo diseño fue realizado para evaluar el desarrollo de enfermedades cardiovasculares con relación a distintos factores de riesgo.

Como ventajas presentan el estar libres de problemas respecto a la selección de testigos y a la precisión de la exposición al agente etológico considerado.

Como inconvenientes están: su alto coste, su largo tiempo de estudio y su exclusiva aplicabilidad a enfermedades relativamente comunes con período de latencia no muy largo.

Asociación y causalidad

Una relación entre dos factores puede ser, falsa, casual o causal.

Las falsas asociaciones suelen estar motivadas por diseños defectuosos que nos conducen a conclusiones erróneas y pueden ser evitadas con una correcta planificación. La asociación casual 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 (según el Diccionario de la Real Academia Española “acontecimiento no previsto cuya causa se desconoce”), 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 la 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 si solos para garantizarnos una relación causal (podría si el estudio fuera aleatorizado, pero no es posible éticamente). Ante esta limitación contamos con estudios adicionales sistemáticos que determinan otras variables para reforzar la hipótesis.

Criterios adicionales en que se basa la relación causa-efecto

El estudio de relación causa-efecto se basa fundamentalmente en descartar explicaciones alternativas, como po-

dría 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 enumeramos a continuación:

1. Fuerza de asociación. Cuanto mayor sea el riesgo relativo o la medida de asociación que utilicemos, mayor será la probabilidad de que el factor sea causa del efecto. Se trataría de estudiar la relación dosis-respuesta que debería ser proporcional: un fumador de 20 cigarrillos/día tiene mayor probabilidad de desarrollar un cáncer de pulmón que otro fumador de cinco cigarrillos, y éste a su vez menos que un no fumador.

2. Plausibilidad y coherencia biológica de la asociación. Los resultados obtenidos son compatibles con los conocimientos anteriores sobre la historia natural y biológica de la enfermedad. Como ejemplo nos puede ilustrar la relación entre el tabaco y el cáncer de cérvix, desestimada en un principio por falta de plausibilidad biológica. Más tarde dicha asociación ha ganado credibilidad al encontrarse en el moco cervical componentes del tabaco. Igual ocurre con la asociación tabaco-cáncer de pulmón.

3. 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, pensando el que método se ha usado y la calidad de cada estudio.

4. Relación temporal entre el factor de estudio y el efecto. Que la causa preceda al efecto. Aunque este principio sea evidente en los estudios transversales y en los de casos y controles con casos prevalentes, esta condición no siempre se puede asegurar.

5. Analogía. Una relación causa-efecto sale fortalecida cuando existe una relación similar en otra situación.

Causas necesarias y suficientes

Una causa es necesaria cuando el fenómeno aparece sólo en su presencia. Por el contrario, es suficiente cuando su sola presencia basta para producir el fenómeno; por ejemplo, una hemorragia grave es causa suficiente de shock, pero no constituye una causa necesaria de muerte, ya que pueden aparecer cuadros de shock sin hemorragia previa (shock séptico o cardiogénico).

En medicina probar ambas condiciones a la vez resulta difícil, debido a la multiplicidad de factores que intervienen e influyen en cualquier fenómeno biológico.

Como norma intentaremos detectar las causas necesarias cuando queramos eliminar un efecto indeseable (fibroplasia retrolenticular en el recién nacido por altas concentraciones de oxígeno inspirado). Por contra, trataremos de detectar las causas suficientes cuando busquemos un efecto beneficioso para el paciente (las vacunas).

Incidencia y prevalencia

1. Incidencia. Se considera incidencia al número de casos nuevos que se producen en una colectividad durante

un período de tiempo determinado. Se calcula a partir de los estudios prospectivos en los que se realiza un seguimiento de los individuos. Hay dos tipos de medida de incidencia:

a) Incidencia acumulada. Sería la proporción de individuos sanos que a lo largo de un período de tiempo desarrollan la enfermedad, siendo necesario especificar el período de observación ya que sin éste no se podría interpretar. Se trata de una proporción más que de una tasa. Su fórmula es:

$$\text{Incidencia acumulada} = \frac{\text{número de casos nuevos en un período de tiempo}}{\text{total de población de riesgo}}$$

b) Densidad de incidencia. Sería el número de casos estudiado en un período determinado de pacientes que desarrollan la enfermedad relacionado con un tiempo determinado de observación. Se utiliza para evitar pérdidas, en los seguimientos de series de pacientes, información sobre ellos y conseguir igualar la duración del tiempo de observación. Es, por tanto, una estimación más precisa y que utiliza toda la información sobre el seguimiento. Su fórmula es:

$$\text{Densidad de incidencia} = \frac{\text{número de casos nuevos en un período determinado}}{\text{total de personas-tiempo en observación}}$$

2. Prevalencia. Es el número de casos existentes de una enfermedad, en un momento dado, en una colectividad. Esta definición corresponde a lo que se conoce como prevalencia de punto o prevalencia a secas. Si se desea prolongar el lapso de tiempo se denomina prevalencia de período, que se define como el número total de casos de la enfermedad durante ese período, esto es, la suma de los casos existentes en la fecha inicial del intervalo del estudio (prevalencia de punto al comienzo del período) más los casos nuevos (incidencia) producidos a lo largo de la duración del tiempo de estudio. Al ser medidas absolutas, son de escaso valor comparativo. Típicamente su estimación se realiza a través de los estudios transversales. Por último, es importante comentar que es por definición una proporción y, aunque coloquialmente nos refiramos a ella como tasa de prevalencia, en realidad no lo es.

Nuevos tests diagnósticos

Toda valoración de nuevos tests diagnósticos se debe realizar mediante el método epidemiológico. Estas mediciones se podrán expresar bajo diferentes aspectos como son: simplicidad (economizando en los costes), objetividad (que puedan ser reproductibles), validez (sensibilidad y especificidad) y aceptabilidad por parte de la población.

Cualquier método diagnóstico puede tener diversas aplicaciones, entre ellas podemos destacar:

1. Cribado. Sospechar la presencia de la enfermedad en la población general, que presenta una baja prevalencia de la enfermedad considerada.

2. Diagnóstico. Para ser aplicado en aquellos que se sospecha padecen la enfermedad. En este caso, la prevalencia de la enfermedad en la población es considerada más alta que en el caso anterior.

3. Selección. Dirigir a los pacientes a otras pruebas más definitivas.

4. Clasificación. Evaluar si una determinada enfermedad es operable o inoperable.

5. Monitorización de una terapia. Actividad de protrombina para los anticoagulantes orales.

Si un investigador desea evaluar una prueba diagnóstica debe meditar con cuidado la aplicación que desea darle, debido a la gran importancia que este aspecto tiene a la hora de seleccionar la muestra a fin de que ésta sea la adecuada para su estudio. Por ejemplo, si un nuevo test se quiere utilizar como cribado diagnóstico en la población general, la muestra debe tener una baja prevalencia de la enfermedad en cuestión. Si por el contrario lo que deseamos es utilizar un método para valorar la presencia de la enfermedad en pacientes en los que se sospecha previamente su existencia, la muestra elegida debe tener una alta incidencia de la enfermedad estudiada. Como norma general, la muestra debe ser extraída del tipo de población al que se desea aplicar el test en la práctica diaria.

Utilizar de modo sistemático muestras con alta prevalencia aumenta artificialmente el valor diagnóstico de un test, el cual posteriormente no resulta ser tan bueno como se pensaba al aplicarlo en la clínica. La sensibilidad y especificidad de la prueba no se ven afectadas por la prevalencia ya que éstos son parámetros intrínsecos a ella. Sin embargo, hace variar sus valores predictivos, sobre todo el positivo.

Grado de enfermedad

Cuando seleccionemos una muestra es importante conocer el grado de enfermedad que presentan los pacientes elegidos. Si la selección se realiza basándose en casos graves y sujetos normales, aumentaremos el valor diagnóstico del test, que posteriormente será poco útil para la amplia gama de pacientes que podemos encontrar en la práctica diaria. Con esto debemos tener siempre en cuenta que la excesiva utilización de casos extremos incrementa artificialmente la sensibilidad y especificidad de cualquier prueba diagnóstica.

Validez de un test

El término validez denota el grado en que una variable mide realmente aquello para lo que esté destinada. Cualquier método diagnóstico debe ser válido, debe medir lo que queremos medir. Para objetivar este punto el investigador debe comparar los resultados del nuevo test con los de un patrón diagnóstico o referencia externa, que actúa como control. Esta referencia es la denominada validación externa, refiriéndose al grado en que los resultados de un estudio pueden ser generalizados a otras poblaciones distintas.

La primera condición que deben poseer unos resultados para poder ser generalizados es que tengan validez interna, es decir, el grado en que los resultados son válidos (libres de error) para la población que ha sido estudiada. La

generalización no se debe hacer a expensas de un solo estudio sino que debemos valorar hipótesis similares estudiadas en poblaciones distintas en ámbitos y zonas geográficas, comprobando que éstas den resultados iguales. También es importante apoyarnos en criterios de plausibilidad biológica (vistos en el artículo anterior), que nos puedan explicar los mecanismos productores de los fenómenos estudiados.

Para considerar una referencia externa óptima debemos escoger la más aceptada al respecto (*gold standard*). Esta referencia nos dará resultados positivos y negativos, tanto verdaderos como falsos. Estos datos nos permitirán con posterioridad calcular la sensibilidad, especificidad y los valores predictivos del nuevo test diagnóstico.

La valoración del test y del método estándar debe realizarse de modo independiente y ciego. Es decir, el investigador que valore el nuevo test debe ignorar los resultados del patrón o referencia externa hasta el momento de realizar la comparación entre ambos.

Evaluación de la validez

Junto con la validez interna y externa se consideran otros tipos de validez que evalúan de forma cualitativa dicha validez, éstos son:

1. Validez de construcción. Sería el grado en que la medición se corresponde con los conceptos teóricos referentes al fenómeno a estudiar.

2. Validez del contenido. Se define como el grado en que la medición abarca la mayoría de las dimensiones del concepto a estudiar.

3. Validez de criterio. Grado en que la medición se correlaciona con un criterio externo de referencia.

El análisis de la validez de criterio para variables dicotómicas se realiza mediante unos índices de validez; los más usados son:

1. Sensibilidad (S). Es un test que mide su capacidad para identificar de manera exacta a los sujetos que padecen la enfermedad. Se expresa en tantos por ciento, y viene definida por la fórmula:

$$S (\%) = \frac{\text{verdaderos positivos}}{\text{verdaderos positivos} + \text{falsos positivos}} = \frac{a}{a+c}$$

Un test diagnóstico tendrá una sensibilidad del 100% cuando no presente ningún falso negativo.

2. Especificidad (E). La especificidad mide la capacidad de la prueba para identificar con exactitud aquellos sujetos que no padecen la enfermedad. Se expresa también como porcentaje, de acuerdo con la siguiente fórmula.

$$E (\%) = \frac{\text{verdaderos negativos}}{\text{verdaderos negativos} + \text{falsos positivos}} = \frac{d}{d+b}$$

Por tanto, una determinada prueba diagnóstica tendrá una especificidad del 100% cuando no exista ningún falso positivo.

Sensibilidad frente a especificidad

En la mayoría de los tests diagnósticos la línea de demarcación que separa a sanos y a enfermos no está clara, y existe una superposición de datos y territorios comunes para ambos. El fijar esta línea es arbitrario por parte del investigador, influyendo directamente en la sensibilidad y especificidad de la prueba.

Ante un corte muy alejado de los valores normales, la especificidad se aproximará al 100% (apenas habrá falsos positivos). Sin embargo, se penalizará la sensibilidad de la prueba, produciéndose muchos falsos negativos. Si situamos la línea de demarcación próxima a los valores normales tendremos el fenómeno inverso. Esto nos indica que sensibilidad y especificidad guardan una relación inversa y ninguna de ellas puede alcanzar valores del 100% mientras existan superposiciones entre la población sana y enferma respecto a la variable medida. La fijación de la línea de demarcación dependerá de ciertas circunstancias clínicas acerca de la enfermedad que estudiamos. Por lo general, debemos exigir una sensibilidad tan alta como sea posible cuando los falsos negativos sean poco deseables. Por ejemplo, en una enfermedad muy rara, para que no pasen desapercibidos los pocos casos existentes, es imprescindible una alta sensibilidad. De igual modo daremos prioridad a la sensibilidad cuando el proceso patológico es muy grave y posea buena respuesta al tratamiento. Por último en un test que se utilice como cribado en la población normal, la sensibilidad también será preferible a la especificidad.

Por el contrario, buscaremos una alta especificidad en todas aquellas circunstancias en que se consideren peligrosos los falsos positivos. Cualquier test diagnóstico para evaluar la muerte cerebral deberá tener una especificidad del 100%. Una alta especificidad es imprescindible también en aquellas pruebas destinadas a confirmar una sospecha diagnóstica previa, o cuando el test está orientado a la evaluación de una enfermedad cuyo tratamiento no modifique sustancialmente la evolución de la misma.

Rango de normalidad en un test cuantitativo

Si la variable es continua, un buen índice de validez es la media aritmética (m) y la desviación estándar (DE) de la diferencia entre el valor de referencia y la media que se emplee. Cuando ambos difieran considerablemente debemos considerarla sesgada. Lo habitual es fijar este rango como el comprendido entre $m \pm 2 \text{ DE}$, que incluye aproximadamente al 95% de la población. Si los márgenes son estrechos, la nueva medida la podemos utilizar con tranquilidad. Si, por el contrario, son muy amplios la medida carecerá de exactitud.

Esta forma de establecer el rango de normalidad sólo es aplicable en aquellos datos numéricos que posean una distribución normal o gaussiana, no siendo éste el caso cuando existen muchas determinaciones clínicas. Por ejemplo, en un estudio hormonal son más frecuentes las cifras "hiper" que las "hipo". Ante una distribución de estas características, o ante la simple sospecha de que podemos estar en presencia de una distribución no gaussiana

na, es mejor establecer los rangos de normalidad por métodos no paramétricos (test de Wilcoxon o rango con signo, test de Mann-Whitney, test de Friedman, test de Kruskal-Wallis); considerando como datos anormales, por ejemplo, todos aquellos que se encuentren por encima del percentil 97 o por debajo del percentil 3. El cálculo de estos percentiles puede ser obtenido en los tratados clásicos de estadística.

Valor predictivo de una medida

Nos planteamos la situación de que un paciente con un resultado positivo tenga la enfermedad en estudio o si, por el contrario, el resultado es negativo, la probabilidad de que no la tenga. Existen unos indicadores que responden a estas preguntas, éstos son:

Valor predictivo positivo (VPP). Es el valor que tiene una prueba positiva para predecir la existencia de la enfermedad. Se calcula como tanto por ciento de acuerdo con la siguiente fórmula:

$$\text{VPP (\%)} = \frac{\text{verdaderos positivos}}{\text{verdaderos positivos} + \text{falsos positivos}} \times 100 = \frac{a}{a+b} \times 100$$

Valor predictivo negativo (VPN). Es el valor que tiene una prueba negativa para predecir lo no existencia de la enfermedad. Se expresa como porcentaje y se calcula del siguiente modo:

$$\text{VPN (\%)} = \frac{\text{verdaderos negativos}}{\text{verdaderos negativos} + \text{falsos negativos}} \times 100 = \frac{d}{d+c} \times 100$$

Los valores predictivos no sólo dependen de la especificidad y sensibilidad, sino también de la prevalencia de la enfermedad.

Interpretación de los valores predictivos

Para la práctica médica, los valores predictivos son más útiles que la sensibilidad y especificidad del test, ya que se adaptan mejor a los requerimientos de la clínica diaria. Pero, como ya hemos dicho, estos valores varían con relación a la prevalencia de la enfermedad en la muestra estudiada. De este modo un determinado test con un alto VPP puede ser de poca utilidad al aplicarlo a una población cuya incidencia de la enfermedad sea notablemente inferior a la de la muestra utilizada en el trabajo original. Pongamos un ejemplo práctico para el mejor entendimiento del problema:

Vamos a estudiar un nuevo test para evaluar la presencia de la enfermedad denominada X, que nos aporta los siguientes resultados:

$$\begin{aligned} \text{Casos estudiados} &= 120 \\ \text{Resultados positivos del nuevo test} &= 63 \\ \text{Resultados negativos del nuevo test} &= 57 \end{aligned}$$

De los 63 casos con test positivo, sólo presentaban realmente la enfermedad 59 pacientes. Entre 57 casos con test negativo existían 6 pacientes que presentaban la enfermedad X. Por tanto:

$$\begin{aligned} \text{Verdaderos positivos} &= 59 \\ \text{Falsos positivos} &= 4 \\ \text{Verdaderos negativos} &= 51 \\ \text{Falsos negativos} &= 6 \\ \text{Sensibilidad} &= 90\% \\ \text{Especificidad} &= 92\% \\ \text{VPP} &= 93\% \end{aligned}$$

Aplicaremos el nuevo test a otra muestra con una prevalencia distinta para evaluar su comportamiento:

$$\begin{aligned} \text{Número de casos} &= 504 \\ \text{Resultados positivos del nuevo test} &= 94 \\ \text{Resultados negativos del nuevo test} &= 410 \end{aligned}$$

De los 94 casos con test positivo, 64 pacientes presentaban la enfermedad X. Por el contrario, entre los 410 con test negativo existían pacientes que presentaban también dicha enfermedad. Los resultados son ahora los siguientes:

$$\begin{aligned} \text{Verdaderos positivos} &= 64 \\ \text{Falsos positivos} &= 30 \\ \text{Verdaderos negativos} &= 405 \\ \text{Falsos negativos} &= 5 \\ \text{Sensibilidad} &= 92\% \\ \text{Especificidad} &= 92\% \\ \text{VPP} &= 68\% \end{aligned}$$

Vemos cómo la sensibilidad y la especificidad apenas se han modificado, ya que son parámetros internos al test. Sin embargo el VPP ha sufrido una importante disminución, que se justifica por la distinta prevalencia de la enfermedad en las dos muestras estudiadas. En el primer caso solo 65 pacientes presentaban realmente la enfermedad X, de un total de 120 casos (prevalencia = 54%). En el segundo supuesto, por el contrario, sólo 69 pacientes presentaban la enfermedad en cuestión de una muestra de 504 casos (prevalencia = 13%). Aunque los valores predictivos dependen de la sensibilidad y especificidad de la prueba, están influenciados de un modo muy especial por la prevalencia de la enfermedad en la muestra seleccionada.

Por último, es importante hacer notar que existe una relación inversa entre VPP y VPN. Teniendo este hecho en cuenta el investigador debe perseguir un VPP alto cuando la finalidad de la prueba diagnóstica sea confirmar la existencia de la enfermedad. Por contra, es deseable un VPN elevado en los casos en que el test vaya destinado a descartar la existencia de una patología concreta.

Investigaciones terapéuticas

Existen diversos trabajos de investigación clínica cuyo objetivo es la evaluación de diferentes terapias. Dichos estudios se suelen realizar en diversas etapas o fases:

– Fase I: descripción de la nueva técnica quirúrgica o dosificación de un fármaco, con evaluación de su toxicidad y efectos colaterales.

– Fase II: investigaciones iniciales para demostrar su eficacia.

– Fase III: estudio controlado frente al tratamiento estándar, para evaluar la eficiencia de la nueva terapia.

Por lo general las dos primeras fases se realizan mediante diseños no controlados, esto es, por medio de estudios descriptivos o de controles históricos. La fase III, por contra, se suele realizar mediante diseños experimentales puros, prospectivos, controlados y aleatorizados.

Los ensayos clínicos destinados a evaluar un fármaco se suelen planificar en 5 fases, a diferencia de los anteriores que los hacen en tres. La primera fase se denomina fase 0 y se realiza en animales de experimentación, se sigue de una fase I cuyo objetivo es establecer la inocuidad del fármaco en voluntarios sanos. Seguidamente se pasará a las tres fases que expuse anteriormente, en este caso serán fase II, III y IV. En esta última se utilizarán preparados ya registrados y comercializados.

El paso por todas estas fases no es obligado para todas las investigaciones, debido a que algunas deberán ser abandonadas si se observa una alta morbilidad o una eficacia dudosa. Por el contrario, habrá circunstancias en las que se podrá demostrar una efectividad muy elevada de un tratamiento nuevo que no justifique desde un punto de vista ético el continuar con un estudio prospectivo, controlado y aleatorizado frente al tratamiento estándar; de este modo condenaríamos a un grupo de pacientes a la peor terapia.

Esquema general de una investigación terapéutica

La base de las investigaciones terapéuticas radica en la aplicación de un tratamiento y el control de la respuesta del individuo a dicho tratamiento. A la nueva terapia la denominaremos tratamiento experimental, aunque otros autores lo definan como intervención experimental o estímulo experimental. Dicho estímulo presenta una intensidad la cual está a veces indicado definirla, como por ejemplo:

1. Dosis y frecuencia de administración de un fármaco o de un tratamiento radioterápico.

2. Intensidad, frecuencia y duración de un estímulo eléctrico.

3. Extensión de una resección quirúrgica.

Se trata, pues, del factor que modificamos con el fin de curar, aliviar sufrimientos, prolongar la vida, mejorar la función o incrementar la calidad de vida. Los distintos tratamientos experimentales pueden ser clasificado del siguiente modo:

1. Cruentos:

– Técnicas quirúrgicas distintas.

– Indicaciones quirúrgicas diferentes.

– Nuevas técnicas anestésicas.

– Distintos cuidados pre, intra o postoperatorios.

– Nuevas prótesis.

2. Incruentos:

– Radiaciones.

– Rehabilitación.

– Dieta.

– Psicoterapia.

– Fisioterapia.

En los estudios controlados de última fase se comparan los tratamientos experimentales frente al control. Por ello el investigador tiene la obligación de ser muy cuidadoso al seleccionar el tratamiento que se aplique a dicho control. Este tratamiento deberá ser el más aceptado y utilizado al respecto.

Tratamiento placebo

El tratamiento experimental tiene con cierta frecuencia un efecto placebo. Definiremos efecto placebo como el efecto psicológico o fisiológico que puede presentar cualquier medicación con independencia del efecto farmacológico que posea. En este efecto intervienen:

– La personalidad del paciente.

– El entusiasmo y el convencimiento del equipo investigador.

– Las características de la intervención y las condiciones de la administración.

Entendemos por tratamiento placebo una intervención de características fisicoquímicas (color, aspecto, calor, olor, etc.) y una pauta de administración similares a las del fármaco a evaluar, pero sin la actividad farmacológica específica para la enfermedad a estudio. En el tratamiento placebo deberemos tener en cuenta las siguientes normas:

1. El dejar sin tratamiento al grupo control no deberá plantearnos problemas éticos especiales. Por ejemplo, aquellas enfermedades graves que no tengan tratamiento eficaz conocido, o bien en enfermedades leves, aunque se disponga de una terapia razonablemente efectiva. El tratamiento placebo está universalmente aceptado en la evaluación de los nuevos fármacos analgésicos, pero sería inaceptable en la valoración de un fármaco antihipertensivo.

2. Con la premisa anterior cumplida, cualquier fármaco nuevo deberá demostrar su eficacia frente a un placebo antes que su eficiencia frente a otros fármacos similares. Esto es, lo primero a demostrar con un nuevo medicamento es que su administración es mejor que no hacer nada. Así pues, si no existen problemas éticos que lo impidan, es aconsejable usar placebo en las primeras fases de la evaluación clínica de los nuevos fármacos, debido a que ciertos aspectos del llamado “ritual médico” pueden mejorar los resultados de la terapia testada. Es muy frecuente encontrar pacientes que se sienten mejor por ingresar en un hospital y percibir que alguien se preocupa por ellos.

3. Estas normas nunca pueden ser aplicadas a tratamientos quirúrgicos. A pesar del efecto placebo indudable de las intervenciones quirúrgicas, la operación simulada *sham operation* es éticamente inaceptable.

La principal ventaja del tratamiento placebo es la de poder controlar los efectos derivados de cualquier característica del tratamiento que se está estudiando. Los principales inconvenientes son éticos, debiéndose utilizar, como dijimos, el tratamiento que haya demostrado mayor eficacia.

También debemos tener en cuenta los inconvenientes prácticos debido a que muchas intervenciones no son susceptibles de ser comparadas con un placebo.

Respuestas al tratamiento

Las respuestas son las variables de desenlace o los criterios de evolución que vamos a utilizar para conocer los efectos de la terapia utilizada. En los ensayos clínicos controlados las respuestas son comunes al grupo control y al grupo experimental.

Las repuestas al tratamiento pueden ser clasificadas del siguiente modo:

1. Clínicas:

- Mortalidad.
- Morbilidad.

2. Laboratorio: Distintas determinaciones analíticas, microbiológicas o anatomo-patológicas.

3. Sociopersonales. Grados de incapacidad, incomodidad o insatisfacción del paciente.

Las normas para cumplir la objetividad y verificabilidad del estudio son las siguientes:

1. Los criterios de evaluación deben ser definidos sin ambigüedades, de la forma más objetiva posible, y antes de iniciar el estudio. Cuantos más juicios tenga que hacer el observador mayor será la tendencia al error experimental.

2. Las respuestas deben tener un sistema de medida que sea objetivo, preciso, válido y fiable. De existir varios métodos, debemos elegir, en lo posible, el más aceptado al respecto.

Respecto al número de respuestas, el investigador ayudado por el entusiasmo tiende a evaluarlo todo. Pero esto no deberemos hacerlo, así pues se deben evaluar tantas respuestas como se crean necesarias, pero tan pocas como sean posibles. La razón de esta afirmación radica en que para cada respuesta deberemos realizar un test de significación estadística de diferencia entre grupos. Por tanto, si evaluamos muchas respuestas, tendremos que realizar muchos tests de significación y aumentaremos las probabilidades de que un test nos salga significativo por azar. De aquí que debamos escoger una o dos respuestas muy relevantes. Lo difícil es preguntarse ¿cuáles serán la más indicadas para cada estudio? Lógicamente variarán según el estímulo que estamos estudiando, si bien nos será muy útil la información bibliográfica. Podríamos enumerar las principales premisas que nos llevarían a escoger las mejores respuestas, éstas son:

1. Deberán ser importantes para el paciente. No debemos incluir complicaciones leves, cuya finalidad sea aumentar la

posibilidad de aumentar la discusión del trabajo. Una conclusión con apariencia de escasa importancia personal puede tener una enorme relevancia social o para la sanidad pública.

2. La respuesta debe ser siempre de aparición frecuente, así evitaremos usar muestras muy grandes.

3. Las respuestas deben ser relevantes, esto es, deben estar relacionadas con la terapéutica estudiada. Deberán tener una alta probabilidad de estar influenciadas por ellas, y caso de no ocurrir así nos veremos obligados a introducir un factor equívoco de nefastas consecuencias para nuestra investigación.

4. Debemos elegir, siempre que nos sea posible, datos fuertes o datos duros, esto es, respuestas cuantitativas (numéricas) que no se presten a interpretaciones más o menos subjetivas por parte del observador o del paciente (cifras de mortalidad, tasas de reoperación, hemorragias postoperatorias, valor hematocrito, presión arterial, diuresis, temperatura, etc.).

Si queremos estudiar un dato débil o dato blando, debido a que las variables cuantitativas no sean las más relevantes para el estudio a realizar, hemos de hacer un esfuerzo por estandarizarlo todo lo posible.

Dentro de los datos débiles escogeríamos variables semi-cuantitativas, objetivadas mediante escalas o *scores*. Hoy día, se dispone de un gran número de ellas que miden dolor, intensidad de un soplo cardíaco, grado de disnea, incomodidad, insatisfacción o bienestar de un paciente para valorar la calidad de vida.

Contamos con variables cualitativas, también llamadas respuestas cuantales o dicotómicas (sí-no, mejor-peor, más-menos, positivo-negativo), caso de no poder disponer de escalas. De esta forma mediremos signos físicos, resultados de múltiples pruebas de laboratorio etc. Al ser menos objetivos los datos obtenidos por este proceder deberemos exigir el consenso entre dos o tres observadores ajenos al estudio, actuando de modo independiente.

Estudios ciegos

Para evitar que tanto las expectativas del médico como las del paciente puedan influir en la evolución de la respuesta observada, se utiliza la técnica de ciego o de enmascaramiento con lo que evitaremos que los implicados en los estudios ante el conocimiento de algunos hechos u observaciones puedan sesgar los resultados.

Existen tres tipos de enmascaramiento:

– *Ciego simple*. Si la valoración de las respuestas es subjetiva por parte del paciente. De esta forma el sujeto no conoce la terapia que recibe, es decir, ignora si pertenece al grupo control o al grupo experimental.

– *Ciego doble*. Cuando la evaluación de la respuesta sea subjetiva por parte del observador. De esta forma tanto el médico como el paciente ignoran la terapia que se ha aplicado en cada momento.

– *Triple ciego*. Cuando otras personas relacionadas con la experimentación (estadístico que analizará los resultados,

personas responsables de decidir la suspensión del tratamiento en caso de reacciones adversas), desconocen el tratamiento que recibe cada paciente.

Lógicamente, los estudios ciegos no son aplicables cuando los efectos de la medicación son obvios y conocidos, ni en aquellos trabajos destinados a comparar los resultados de un tratamiento médico contra un tratamiento quirúrgico.

El estudio doble ciego es la única forma de garantizar una observación objetiva, sin ideas preconcebidas del investigador ni ayudas bienintencionadas al paciente. Sin embargo, no estamos de acuerdo con algunos autores que exigen el doble ciego de forma sistemática, aplicando este diseño a toda clase de estudios en los que sea viable, y de forma muy particular a todos los ensayos clínicos de nuevos fármacos. Ante unas respuestas a evaluar objetivas, y que no lleven consigo ningún juicio de valor por parte del investigador, la precaución de realizar un diseño doble ciego será probablemente innecesaria.

Cumplimiento de la terapia

El cumplimiento riguroso de una terapia no es fácil, sobre todo si los tratamientos son largos, intensos o complejos.

No cumplir la terapia dificulta mucho la interpretación de los resultados obtenidos. Si se incluyen los malos cumplidores estamos analizando un subgrupo de pacientes que difícilmente han podido beneficiarse de una terapia defectuosa o inexistente. Por contra, el excluir a los que no hayan cumplido las normas compromete la validez interna del estudio, debido a que la falta de cumplimiento pudiera estar relacionada con el tipo de tratamiento prescrito (complejidad, efectos colaterales).

En ambos casos será difícil solucionar el problema y determinar la influencia del mal cumplimiento en los resultados obtenidos. Para ello debemos extremar las medidas para que las terapias se cumplan dentro de lo razonable.

Por último, excluiremos a todos aquellos paciente que probablemente no sean capaces de seguir de forma escrupulosa y disciplinada el protocolo. Ante la duda recurriremos a la llamada fase de adaptación, previa al estudio. Con ella podremos excluir a los pacientes que no hayan observado con rigor las medidas prescritas. Es nuestro deber, una vez comenzado el estudio, la mentalización de los pacientes sobre la importancia del cumplimiento estricto de la terapia, intentando establecer algún método para controlar dicho cumplimiento (contar las cápsulas, por ejemplo).

Población y muestra. Teoría del muestreo

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 denominamos muestra. Los resultados obtenidos con esa muestra serán los que posteriormente se generalicen extrapolándolos a toda la población mediante cálculos estadísticos apropiados.

Población

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. 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.

– Población de estudio accesible. Es 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.

Definición de población o participantes

Para conseguir una muestra homogénea, el investigador utiliza la definición de la población y los criterios de inclusión y exclusión.

1. Criterios de Inclusión. 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 (clínicos, analíticos, radiológicos), debiendo el investigador elegir los más aceptados al respecto.

Para ello nos es de gran utilidad la bibliografía. No obstante es importante tener en cuenta algunas recomendaciones:

– Siempre que ello sea posible deberemos escoger criterios cuantitativos (numéricos): cifras de tensión arterial, valores analíticos, tallas, pesos.

– Debemos evitar los límites entre lo normal y lo patológico, con ello no incluiremos individuos sanos en nuestro estudio, y tampoco a los especialmente graves.

– Si es necesario, la utilización de criterios cualitativos (dolor, palidez) o semicuantitativos (intensidad de un soplo cardíaco), sería importante buscar un consenso entre dos observadores neutrales, ajenos al estudio y al equipo investigador, caso de necesidad por desacuerdo se podría recurrir a un tercer observador.

– Es importante utilizar, siempre que sea posible, más de un criterio para definir la enfermedad.

Los criterios de inclusión deberán quedar establecidos de un modo claro, objetivo y preciso, para evitar posteriores dudas al respecto.

2. Criterios de exclusión. Algunos pacientes que hayan sido definidos previamente deberán ser excluidos del estudio, sobre la base de dos consideraciones fundamentales:

– Por representar condiciones que supongan un riesgo adicional para ese paciente (viejos, niños, mujeres gestantes, pacientes con determinadas alergias, etc.).

– Por presentar factores 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.

Homogeneidad y representatividad de la muestra

Mediante la definición de la población, el investigador delimitará el tipo de pacientes que va a estudiar, pero también está obligado a definir qué signos o síntomas marcarán la enfermedad a estudiar, y de esta forma conoceremos la representatividad de la muestra.

Como sabemos, una de las características 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 a 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 podrían explicar por sí sola las diferencias encontradas en el estudio.

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 pacientes 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.

Teoría de muestreo

Se define "muestreo" o extracción de muestras como el 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, pudiendo a veces una muestra generar datos más exactos que toda la población de donde se extrajo, lo que nos hace pensar que los errores provienen no sólo de los errores de muestreo, sino también de las inadecuaciones de dicha tarea.

De lo anterior se deduce que el investigador rara vez podrá incluir en un estudio a todos los pacientes disponibles que cumplan las condiciones basales, caso de hacerlo estaríamos realizando un "estudio abierto", destinado a investigar patologías poco frecuentes. Sería por tanto lo normal realizar muestras extraídas de la población estudiada.

Técnicas de muestreo

Para que una muestra sea representativa de la población a estudio debemos garantizar que todos los sujetos integrantes de dicha población han tenido igual probabilidad de entrar a formar parte de la investigación.

Los planes de muestreo pueden dividirse en dos grandes grupos:

Muestreo probabilístico. Es aquel que utiliza formas de selección aleatoria para escoger unidades idóneas. La característica fundamental de la muestra probabilística es que el investigador puede especificar la probabilidad de que se incluya en la muestra cada elemento de la población.

Es importante conocer dos conceptos dentro del muestreo probabilístico; éstos son:

– Unidad de muestreo. Es el elemento sobre el que se aplicará la técnica de selección (personas, centros sanitarios, escuelas, etc.).

– Fracción de muestreo. Es el cociente entre el tamaño de la muestra deseado y el total de la población candidata.

Los cuatro métodos más utilizados para la extracción de muestras probabilísticas son:

1. Muestreo aleatorio simple. Es el método que mejor garantiza la representatividad de la muestra, pero exige conocer la lista completa de la población objeto de estudio. Esto lo convierte en una técnica extremadamente laboriosa, no siendo siempre posible de realizar, de ahí que se use tan poco en la práctica. En este tipo de muestreo los pacientes se eligen mediante sorteo interviniendo sólo el azar puro, lo que puede conducir a muestras altamente sesgadas respecto a la población, especialmente si se trata de muestras pequeñas.

2. Muestreo aleatorio estratificado. Es una variedad del muestreo aleatorio simple en la que se divide en primer lugar a la población en dos o más estratos o subgrupos. Con ello intentamos lograr un mayor grado de representatividad y evitar la posibilidad de obtener muestras sesgadas al conseguir una igualdad en la distribución de variables.

La estratificación puede estar basada en atributos como sexo, edad, ocupación laboral, etc. La principal ventaja estriba en obtener estimaciones más precisas de los parámetros poblacionales que con el muestreo aleatorio simple, ya que la variancia total se basa en la de cada estrato o subgrupo y al ser éstos más homogéneos que la población total, tienen menor variabilidad.

El principal problema de esta técnica radica en la dificultad para identificar las variables de mayor interés.

3. Muestreo por cúmulos. Hay situaciones en las que es imposible realizar un listado de todos los elementos de

una población con el fin de extraer de ella una muestra simple o estratificada de tipo aleatorio. En esta circunstancia contamos con el muestreo por cúmulos en el que se realiza una obtención aleatoria sucesiva de unidades. Así pues, la primera unidad a extraer es un cúmulo o grupo grande (unidad primaria), posteriormente se extraerá, en una segunda etapa, una muestra de cada una de las unidades primarias (unidades secundarias). Podemos utilizar el número de etapas que creamos convenientes y en cada una de ellas aplicaremos un método distinto de muestreo (simple, estratificado, etc.). Esta técnica también se conoce como “muestreo de múltiples etapas”.

En el muestreo de cúmulos o de múltiples etapas, la variabilidad es mayor que en el caso del aleatorio simple o el estratificado, y su cálculo más complejo. Una solución pasa por extraer una muestra de mayor tamaño y considerarla como aleatoria simple, intentando que las unidades de muestreo sean lo más pequeñas posibles.

Su principal desventaja es que caso de que las unidades primarias contengan personas similares con relación al fenómeno de interés, será difícil estimar la precisión con la que pueden calcularse los parámetros poblacionales.

4. Muestreo sistemático. Caso de que el investigador no conozca la lista completa de los pacientes potenciales elegibles puede optar por un criterio más o menos sistemático, en función de un determinado período designado por azar (uno de cada diez pacientes que acuden a consulta) o del número de la historia clínica (los acabados en cero). Lo importante en este caso es garantizar que la constante no guarda relación con la variable a medir.

La principal ventaja es la facilidad de aplicación, lo que conlleva a cometer menor número de errores, de igual modo no siempre será necesario tener con antelación una lista completa de toda la población; por último, nos beneficiaremos caso de que la población de referencia esté ordenada siguiendo una tendencia conocida (de viejo a joven, de alto a bajo, etc.) ya que el muestreo sistemático asegura una cobertura de unidades de todos los tipos.

Muestreo no probabilístico. Es una técnica que comprende la selección de una muestra por lo regular tenida en menor estima que la que obtenemos por muestreo probabilístico, no obstante innumerables muestras obtenidas en investigación no son probabilísticas.

Se conocen tres métodos de muestreo no probabilístico:

1. Muestreo accidental. Conlleva el uso de las personas u objetos, con los que fácilmente se cuenta, como sujetos del estudio, aunque no tienen que estar obligadamente integrado por personas que conozcan los investigadores.

El principal problema de estas muestras es que los sujetos accesibles puede que no sean típicos de la población, con relación a las variables críticas que se intentan medir.

Otro tipo de muestreo accidental es el denominado de “recomendación”, que se utiliza cuando las poblaciones en estudio comprenden personas con rasgos específicos que son difíciles de identificar por medios ordinarios.

Aunque es la técnica menos consistente de esta categoría es la más utilizada, si bien recomendamos al investigador extreme las precauciones al analizar e interpretar los datos resultantes de ellas, caso de utilizarlas.

2. Muestreo por cuotas. En esta técnica el investigador utiliza algunos conocimientos sobre la población para impartir alguna representatividad al plan de muestreo. Así pues la composición general de la población de referencia se realiza por características tales como sexo, edad, etc., y es conocida de antemano. El principal requerimiento es conseguir el número adecuado de personas para llenar cada una de las cuotas escogidas. Desde el punto de vista técnico, aunque mejor, es muy semejante al muestreo accidental, debiéndose considerar su uso si los recursos impiden la utilización de un plan de muestreo probabilístico.

3. Muestreo consecutivo. Se trata de incluir a todos los pacientes atendidos de modo consecutivo durante un período de tiempo concreto y determinado.

Se deben buscar períodos naturales (de enero a enero), evitando en lo posible utilizar “número prefijados” (“a propósito de 100 casos”) que nos haría preguntar qué ocurrió con el caso ciento uno. Es la técnica más utilizada en los ensayos clínicos especialmente si se trata de enfermedades agudas.

4. Selección de voluntarios. Esta técnica es sospechosa de sesgos debido a que el paciente que acude voluntariamente a un estudio puede presentar algunas características especialmente físicas o psicológicas que justifiquen su decisión, lo que afectaría a la representatividad de la muestra.

Tamaño de la muestra

Todo investigador debe estudiar un número suficiente de casos para obtener resultados clínicamente relevantes, pero no tanto como para despilfarrar 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 puede depender del grado de precisión requerido, lo heterogéneo de los atributos en investigación, las diversas técnicas de muestreo que utilizaremos, y el tipo de técnica de estimación usado. Volvemos a necesitar en este apartado la bibliografía para orientarnos en este sentido. Si la bibliografía no nos saca de dudas podremos consultar con un estadístico que nos calculará qué tamaño debe tener nuestra muestra en función de la variabilidad del parámetro que queramos estudiar, así como la precisión de la estimación y el nivel de confianza deseado.

En otras ocasiones tendremos que recurrir al llamado “estudio piloto” o preliminar, realizado a base de una pequeña muestra.

Los resultados obtenidos serán estudiados por un estadístico, calculándose el tamaño aproximado de la muestra del estudio definitivo.

Planteamos las siguientes consideraciones generales a tener en cuenta:

1. Es aconsejable huir de las muestras pequeñas, y no estudiar grupos de menos de 10 individuos. Con ello nos evitaremos que la homogeneidad de una muestra semejante quede en entredicho y la razonable dificultad del análisis estadístico.

2. Si se trata de un estudio controlado, ambos grupos deberán ser similares respecto a número de pacientes (relación 1:1). Se podría aceptar hasta una proporción 1:3 sin grandes problemas, pero diferencias mayores llevarán a que el análisis estadístico sea menos fiable.

3. Si el investigador pretende demostrar ausencia de diferencias entre dos grupos de estudio, el cálculo de la muestra es de sumo interés. Así pues si queremos demostrar que un tratamiento A y B tienen unos resultados similares, debemos estar muy seguros de que el tamaño de muestra es suficiente para detectar la diferencia, si la hubiera.

4. Por último hacer notar que la muestra debe quedar determinada desde el principio o después del estudio piloto. No debemos dejarnos influenciar por los resultados parciales que se obtengan ni realizar pruebas repetidas de significación estadística con este fin. Esto nos llevaría a incrementar la probabilidad de obtener una diferencia significativa por puro azar, no reflejando ésta un comportamiento realmente distinto de ambos grupos de estudio.

Estudios multicéntricos

Debido a la modestia cuantitativa de los avances en medicina clínica, la investigación necesita de estudios experimentales muy potentes para detectarlos. Caso de no disponer de un número de pacientes suficientes y requerir de una muestra grande, podemos recurrir a algunos de los procedimientos que enumeramos a continuación:

1. Búsqueda de técnicas más precisas para medir las variables a estudio, disminuyendo de esta forma la dispersión de los datos y el tamaño de la muestra.

2. Utilizar diseños apareados, cruzados o estratificados, que aumentan la homogeneidad de los grupos y disminuyen el tamaño de la muestra.

3. Elaboración de criterios de selección menos estrictos, con lo cual podríamos incluir más enfermos, a expensas de penalizar la validez interna del estudio (homogeneidad).

4. Alargar el estudio en el tiempo, permitiendo con ello incluir mayor número de pacientes.

A pesar de todo ello, en determinadas ocasiones nos vemos obligados a recurrir a los denominados "estudios multicéntricos", en los que colaboran varias instituciones. Obtenremos con ellos 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.

– La disminución de la homogeneidad de la muestra, y el aumento de forma paralela la dispersión de los datos.

Para obviar estos inconvenientes deberemos elaborar unos criterios de selección de pacientes muy simples, evitando subjetividades por parte de los distintos investigadores. Junto con esto, se recomienda una buena comunicación entre los distintos grupos de trabajo, condición a veces imprescindible para unificar criterios. El mejor modo de conseguir esta comunicación es con reuniones de trabajo muy frecuentes, labor que debe asumir el director del estudio.

PASOS A DESARROLLAR EN LA EJECUCIÓN DE UN TRABAJO CIENTÍFICO

Una vez elaborado el plan experimental, de acuerdo con lo expuesto anteriormente, estamos en disposición de iniciar la fase práctica de la investigación propiamente dicha. No debemos olvidar lo que todos los investigadores experimentados saben: los trabajos científicos suelen estar regidos por la "ley de la perversidad general o ley de Murphy": si algo puede fallar, fallará, y generalmente en el momento menos oportuno. De aquí la gran importancia de los pasos preliminares a los que el investigador deberá prestar atención antes de pasar a la fase de ejecución. Dichos pasos son:

Conversaciones con el estadístico

El estadístico es una pieza clave en todo trabajo de investigación que se precie. La bioestadística nos será imprescindible para el análisis de los datos experimentales, pero también nos aportará ideas de gran valor para el diseño, muestreo y obtención de datos. Por ello se debe consultar al estadístico antes de recoger los datos ya que de hacerlo después, los fallos cometidos no tendrán solución.

Debemos tener en cuenta que los bioestadísticos no suelen saber medicina clínica, lo que conlleva a una limitación en el diálogo con él. De igual modo, el investigador deberá conocer las posibilidades estadísticas de su trabajo (ayudado fundamentalmente por la bibliografía), lo que potenciará un mejor entendimiento entre investigador y bioestadístico. Conseguiremos así un plan experimental, y elaboraremos las normas para la recogida de los datos precisos con el máximo rigor posible.

Formulario de recogida de datos

Los datos generales en la investigación se suelen trasladar a un "formulario", que facilita el análisis posterior por parte del estadístico. Deberá realizarse un formulario por cada paciente, donde quedarán recogidos los datos correspondientes.

Los principios de un buen formulario son:

1. Simplicidad. Cuanto más simple y más breve sea, más fácil será la tarea, por ello los mejores son los de uno o dos folios.

2. Claridad. Conlleva a incluir un determinado número de instrucciones para cumplimentar correctamente el formulario. Si el formulario es complejo será necesaria una lista de chequeo final, para estar seguros de no haber olvidado ningún aspecto importante durante la cumplimentación.

3. Eficacia. El formulario debe ser organizado de forma lógica y eficaz, evitando así pérdidas de tiempo para el que lo rellena. Como norma, se deben apuntar los datos en que se generan.

Todo formulario debe incluir el nombre del paciente, el número que ocupa dentro del estudio y la firma del investigador responsable de su cumplimentación. Es importante dejar un apartado en blanco para observaciones, donde el investigador pueda apuntar algunas circunstancias inusuales del caso que puedan ser de interés al analizar los datos. Para evitar sorpresas se debe intentar que cada formulario sea llenado por una persona y guardarlo en lugar seguro. De ser posible es aconsejable guardar un duplicado (fotocopia) para cada formulario, archivándolo en un lugar distinto del original.

Protocolos de investigación

Todo investigador puede verse en un momento de su vida obligado a elaborar una serie de "memorias, protocolos o proyectos de investigación" con distintos fines, esto es:

- Informar a su director de tesis.
- Pedir consejos a otros investigadores más expertos.
- Inscribir la tesis doctoral, de acuerdo con la legislación vigente.
- Solicitar una ayuda económica para su trabajo.
- Pedir una beca de investigación.
- Obtener del comité ético, de investigación o de ensayos clínicos el permiso correspondiente para iniciar su trabajo.
- Informar al bioestadístico.

Excluyendo las razones administrativas y económicas, el protocolo también es útil como documento interno o manual de operaciones de equipo investigador, para unificar y sistematizar criterios.

El esquema general de un protocolo de investigación sería:

1. Introducción: importancia del problema.

2. Material y métodos:

- a) Definir la enfermedad.
- b) Condiciones basales: criterios de inclusión y exclusión.
- c) Muestra: selección, tamaño, grupo de estudio.
- d) Obtención y manejo de muestras biológicas.
- e) Definición del tratamiento (control y experimental).
- f) Respuestas: número, formas de medirlas, cronología de las medidas.
- g) Tratamiento estadístico de los datos, seguimiento de los pacientes.

3. Bibliografía.

4. Anexos (opcionales): formulario de recogida de datos, modelo de consentimiento informado, etc.

Debemos velar siempre por la claridad y el detalle, asegurándonos que quedan perfectamente definidos los pro-

pósitos y objetivos del estudio (y su importancia potencial, pero sin exagerar), así como los perfiles fundamentales del diseño, muestreo y tratamiento estadístico de los datos. La bibliografía deberá incluir todos los trabajos importantes al respecto, aunque no es necesario ni conveniente que sea exhaustiva. Todo ello deberá ser breve, con máximo de 15 folios, aumentando así las posibilidades de ser leído con atención por las personas a las que va destinado.

Control de calidad de la información

Durante la obtención de los datos hemos de evitar, en lo posible, ciertos errores que nos llevarían a resultados inaceptables o falsos que invalidarían el estudio. Dentro de estos fallos existen algunos que son casuales y por tanto debidos al azar. Éstos son inevitables, y se pueden estimar adecuadamente por los métodos estadísticos habituales. Otros errores, por contra, son sistemáticos, motivados por una observación o medición inadecuada, y deben ser evitados a toda costa.

Para evitar caer en este problema debemos validar nuestros métodos para garantizarnos unos resultados correctos. Caso de que la investigación sea larga, se deberá realizar un control de calidad cada cierto período de tiempo. Esto implica la realización de un pequeño estudio piloto, por medio del cual se intentarán detectar los errores de medida que pudieran existir, imputables al observador o al aparataje utilizado.

Errores propios del observador

Que un dato cuantitativo medido con el método apropiado sea objetivo y confiable depende de lo adecuada que sea nuestra medida. Por ello es aconsejable calcular las llamadas diferencias (o variaciones) inter e intraobservador.

Se denomina variación interobservador aquella diferencia que se produce entre las valoraciones hechas por dos o más observadores expertos actuando de modo independiente, al realizar la misma medida o determinación. Si, por el contrario, es el mismo observador el que realiza idéntica medida en varias ocasiones separadas entre sí por un lapso de tiempo, estaremos calculando la variación intraobservador.

Para evaluar estas variaciones utilizaremos los test estadísticos de diferencias de medias (t de Student o ANOVA, según se trate de dos o más observaciones). También podemos utilizar para el análisis el cálculo del correspondiente coeficiente de correlación entre las distintas observaciones. Caso de existir diferencias significativas en el primer supuesto ($p < 0,05$), o la correlación ser débil en el segundo supuesto (en general, $r < 0,6$), las diferencias encontradas son inaceptables, hemos de estandarizar mejor la medida, y ello nos puede llevar más o menos tiempo, según la complejidad de la prueba. Si no podemos estandarizarla adecuadamente, utilizaremos como dato la media de varias medidas realizadas por el mismo observador (si es la variación intraobservador la afectada), o la media de las medidas realizadas por dos o más observadores distintos e independientes (si era la variación interobservador la defectuosa).

Datos cualitativos

Aunque se podría utilizar un coeficiente de correlación para datos cualitativos ("r" de Spearman), lo normal es calcular las variaciones inter e intraobservador mediante el llamado índice de concordancia, por medio del cual se exige un acuerdo mínimo del 95% de los casos; tanto el mismo observador como observadores diferentes. Si no alcanzamos este valor mínimo procederemos a estandarizar mejor la estimación, o a buscar el acuerdo entre dos o más observadores diferentes actuando de modo independiente.

Errores propios de los instrumentos

Todo aparato utilizado para realizar determinaciones bioquímicas debe estar certificado y calibrado y sus reactivos tendrán un control de calidad adecuado. No obstante corremos el riesgo de que no esté en condiciones óptimas para investigar con él, aunque se utilice en la clínica diaria de forma aceptable.

La verificación de un aparato se lleva a cabo mediante la repetición de la medida con la misma muestra, tanto en el mismo momento (variación intraensayo) como en días distintos (variación interensayo). Hallaremos la media y la desviación estándar de las diferentes medidas, y los correspondientes coeficientes de variación se calcularán mediante la siguiente fórmula:

$$\text{Coeficiente variación (\%)} = \frac{\text{desviación estándar}}{\text{media}} \times 100$$

El coeficiente de variación intraensayo debe ser menor que el interensayo, debido a que correspondiente mediciones se realizan el mismo día, en iguales condiciones y utilizando los mismos reactivos. En cualquier caso, ninguno de los dos debe superar el 5%. Si cualquiera de ambos supera los límites tolerables debemos revisar el aparato y las proporciones utilizadas. Caso de no poder mejorar la calidad de las mediciones utilizaremos como dato la media de varias medias sucesivas, o buscar otro aparato más fiable.

Tratamientos estadísticos de los datos experimentales, interpretación de los resultados y búsqueda bibliográfica.

Encuesta, entrevista y cuestionarios

Podríamos definir "la estadística" como el método de razonamiento que permite interpretar un conjunto de datos cuyo carácter esencial es la variabilidad (variables).

Por ello una vez obtenidos los datos, el investigador tiene la obligación de describir su muestra correctamente a fin de que otros investigadores puedan apreciar sus características. De este modo se podrá saber si las conclusiones pueden ser aplicables al tipo de pacientes que suele tratar en su medio.

El modo más común de realizar una generalización empírica de los resultados en la investigación clínica es mediante la aplicación de la llamada estadística analítica o inferencial. Esta comprende las bases lógicas para establecer conclusiones relacionadas con las poblaciones a partir de los resultados obtenidos, todo ello basado en la teoría probabilística. Aplicarla en la fase de análisis tiene dos fines:

- Evaluar la variación aleatoria.
- Controlar los factores de confusión.

Aunque la mayoría de los médicos se sientan desbordados por la complejidad de la estadística, debemos decir que sus métodos no prueban ni demuestran nada, solo se limitan a calcular el grado de probabilidad (o el margen de error) con el que hacemos la extrapolación antes mencionada.

Por suerte para nosotros no se trata de aprender matemáticas, sino de conocer sus aplicaciones y limitaciones. Con ello podremos valorar correctamente el 90% de los trabajos conociendo poco más de media docena de técnicas estadísticas muy habituales, como son: la t de Student, la χ^2 , límites de confianza, ANOVA, rectas de regresión, curvas de supervivencia U de Mann-Whitney, etc.

La estadística se utiliza para describir y sintetizar datos obtenidos de observaciones y mediciones empíricas. Como ejemplo de estadística descriptiva están los promedios y los porcentajes. Cuando estos índices se calculan en datos de una población, se conocen como parámetros, mientras que el índice descriptivo de una muestra recibe el nombre de estadístico.

No obstante, tras tan largo camino recorrido a través del artículo actual, sería duro para el lector introducirnos de lleno en el tratamiento estadístico de los datos experimentales, la interpretación y presentación escrita de los mismos, la técnica de búsqueda bibliográfica y el modo de hacer las encuestas, entrevistas y cuestionarios. Por ello pienso que dichos temas tienen envergadura suficiente como para ser tratados individualmente en la sección de "Formación Continuada" de nuestra revista.

A modo de conclusión de este artículo, sobre la investigación en atención primaria, me gustaría enviar a todos los lectores un mensaje de optimismo ante este mundo tan apasionante de la investigación. Piensen que aunque se supone que todo está hecho respecto a investigación, quedan lagunas inmensas por llenar en muchos campos de la atención primaria y somos nosotros, los médicos de familia, los llamados a cumplimentar esa faceta investigadora que nuestra calidad asistencial nos reclama.

Desde estas líneas y como vocal nacional de investigación de nuestra Sociedad, estimulo a todos y cada uno de vosotros a continuar, los que ya hayan empezado, y a comenzar los que no lo hayan hecho aún, en este campo cada vez más necesario, prolífico y gratificante de la investigación científica.

BIBLIOGRAFÍA GENERAL

- Aleixandre R, Porcel A, Aguiló A, Marset S, Abad F. Diez años de la revista *Atención Primaria* (1984-1993): análisis bibliométrico y temático. Aten Primaria 1996; 17: 225-230.
- Argimón Pallás JM, Jiménez Villa J. *Métodos de Investigación. Aplicados a la atención primaria*. Madrid: Harcourt Brace, S.A., 1991.
- Argimón Pallás JM, Jiménez Villa J. *Investigación en atención primaria*. En: Martín Zurró A, Cano Pérez JF, editores. *Atención Primaria* (4.ª ed.). Madrid: Harcourt Brace, 1999; 161-182.
- Baladrón J, Vaga D, Aldecoa B. *Manual intensivo para examen MIR*. Madrid: Editorial Lusán, 1991.

- Cañedo L. Investigación clínica. México Ed. Interamericana, 1987.
- Carbó JM, Arnau C, García-Barbal J. La metodología y el paciente. *Aten* 1997; 1223: 80.
- Carrasco de la Peña J. El método estadístico en la investigación médica. Madrid: Ed. Ciencia 3, 1986.
- De la Fuente JA, Mercader J, Borrel C, Martín C, Birulés M, Marcos L et al. Factores relacionados con la actividad investigadora de los médicos de atención primaria. *Aten Primaria* 1997; 19: 283-289.
- Delgado Marroquín MT. Investigación en atención primaria: ¿una cuestión de calidad? *Jano* 1995; 1132: 1464.
- Diccionario Enciclopédico (6.^a ed.). Madrid: Espasa-Calpe, 1990.
- Europa Científica. Investigación y Tecnología en 20 países. Barcelona: Ed. Labor, 1990.
- Fernández de Sanmamed Santos MJ. Introducción a la investigación cualitativa. FMC 1995; 1: 23-28.
- Gelijns AC, editor. Modern methods of clinical investigation. Washington: National Academy Press, 1990.
- Gérvás JJ. Prólogo. En: Howle JGR, editor. Investigación en medicina general. Madrid: Díaz de Santos, 1993.
- Goldstone, LA. Understanding medical statistics. Londres: William Heineman Medical Books Limited, 1983.
- Gómez de la Cámara A. Investigación en atención primaria. MEDIFAM 1992; 2: 63-65.
- Gómez de la Cámara A. La investigación en atención primaria. El ensayo clínico y los estudios observacionales de productos farmacéuticos. *Aten Primaria* 1999; 24: 431-435.
- Gómez-Calcerrada D, Hernández del Pozo F, Gómez Berrocal M. La investigación en la utilización de servicios sanitarios. FMC 1996; 3: 229-242.
- González Egúaras M, García Criado E. ¿En qué consiste el método experimental? SEMER 1992; 6: 449-450.
- González Egúaras M, García Criado E. Pasos a desarrollar para la ejecución de un trabajo científico. SEMER 1994; 2: 101-103.
- Gosalbes Soler V. ¿Qué debemos investigar en atención primaria? Centro de Salud 2000; 8: 492-494.
- Hawkins C, Sorgi M. Investigación médica. Barcelona: Ediciones Médici S.A., 1990.
- Jiménez Villa J. Evolución de la investigación en atención primaria. *Aten Primaria* 1993; 11: 115-116.
- Juncosa Font S. Más y mejor investigación en servicios de atención primaria. MEDIFAM 1998; 8: 13-19.
- Kelsey JL, Thompson WD, Evans A. Methods in observational epidemiology. Nueva York: Oxford University Press, 1986.
- Lázaro P, Pozo F, Ricoy JR. Una estrategia de investigación en el sistema nacional de la salud (II): Investigación en servicios de salud. *Med Clin (Barc)* 1995; 104: 67-76.
- Levine RJ. Ethics and regulation of clinical research (2.^a ed.). Baltimore: Urban and Schvarzenberg Inc, 1986.
- Lloveras Cáñaves J. Reflexionando sobre los obstáculos en la investigación de atención primaria. Opción de los investigadores. En: Documentos de trabajo sobre investigación en Atención Primaria. Madrid: FIS, 1996.
- López-Miras A, Pastor Sánchez R, Gérvás J. Investigación en atención primaria: contenido de los artículos publicados en revistas científicas. *Aten Primaria* 1993; 11: 340-343.
- Malacara JM. Bases para la investigación biomédica. México: Distribuidora y Editorial Mexicana, S.A. de CV, 1987.
- Martín Moreno JM, Guallar E, Rodríguez Artalejo F. La importancia de la elección de la pregunta de investigación. *Jano* 1995; 1148: 1675-1681.
- Martín Zurro A. Investigación en atención primaria. En: Argimón Pallàs JM, Jiménez Villa J, editores. Métodos de Investigación. Aplicados a la atención primaria. Madrid: Harcourt Brace, S.A., 1991; 1-6.
- Martínez JA. La investigación en atención primaria. *El Médico*, Anuario 1997.
- Matthews DE, Farewell VT. Estadística médica: aplicación e interpretación. Barcelona: Salvat, 1988.
- Melguizo M, López MA. La investigación en atención primaria. *Aten Primaria* 1991; 8: 359-360.
- Morton RF, Hebel JR. Bioestadística y epidemiología (2.^a ed.). Madrid: Interamericana-McGraw-Hill, 1987.
- Laín Entralgo P. Historia de la medicina. Barcelona: Salvat Editores, 1978.
- Payán MJ. Los ensayos clínicos en atención primaria ¿Investigación o promoción? *El Médico* 1999; 3: 32-38.
- Pico MV, Delgado M, Altisent R. Aspectos éticos y legales de la investigación en atención primaria. FMC, 1995; 2: 212-219.
- Piédrola G, Domínguez M, Cortina P, Gálvez R, Sierra A, Sáez MC et al. Medicina preventiva y salud pública (8.^a ed.). Barcelona: Salvat, 1990.
- Polit D, Hungler B. Investigación científica en ciencias de la salud (3.^a ed.). México: Interamericana, 1995.
- Pritchard P. Manual de atención primaria de salud. Su naturaleza y organización. Madrid: Díaz de Santos, 1986.
- Ricoy JR, Guasch MF, Clavería LE, editores. Una aproximación al análisis de un Programa de I+D. Madrid: Fondo de Investigación Sanitaria. Ministerio de Sanidad y Consumo, 1996.
- Ríos S. Iniciación estadística. Madrid: Paraninfo, 1992.
- Rodríguez del Águila MM, Fernández Sierra MA. Síntesis de las principales técnicas estadísticas aplicadas en la investigación sanitaria. *Jano* 1998; 1263: 48-52.
- Starfield B. A framework for primary care research. *J Farm Practice* 1996; 42: 181-185.
- Starfield B. Prioridades y futuro de la investigación en atención primaria. *Cuadernos de Gestión*. 1997; 3: 201-210.
- Swinscow TDV. Estadística primer nivel. Barcelona: Salvat, 1990.