

# Capítulo 12:

## Tipos de diseños

### Presentación

Este capítulo describe las herramientas de las que dispone el investigador para controlar las terceras variables: primero los métodos para aquellas que son observables, y luego cómo el azar ayuda a controlar las terceras variables no conocidas o no observables. Finalmente, se describen los diseños más habituales en Ciencias de la Salud.

### Objetivos

#### Al terminar este capítulo, un lector que haya realizado los ejercicios:

- Distinguirá entre variables de las unidades (condiciones Z) y variables del investigador (intervenciones X).
- Se preocupará por observar si el criterio de ajuste se ha decidido independientemente de los resultados.
- Valorará más un ajuste ya preespecificado en el protocolo.
- Sabrá que los principales métodos de ajuste son la restricción, el estudio de subgrupos y el modelado estadístico.
- Se preocupará porque el estudio de subgrupos persiga que la intervención en estudio esté balanceada en todos los grupos.
- Identificará el apareamiento con el caso extremo de subgrupos en el que cada subgrupo está formado por una pareja.
- Sabrá que el modelado también conviene especificarse en el protocolo.
- Interpretará a nivel exploratorio los ajustes no especificados en el protocolo.
- Centrará el interés de un estudio de intervención en el efecto de la causa X en la respuesta Y dadas las condiciones Z.
- Al diseñar estudios confirmatorios de intervención, las relaciones de Z con X y de Z con Y formarán parte de las premisas previas.
- En los estudios exploratorios de intervención, antes de ajustar por Z se preguntará por la viabilidad de intervenir sobre X a nivel fijo de Z.
- En presencia de modificación del efecto, preferirá empezar por el estudio del efecto en un único subgrupo.
- Distinguirá entre las condiciones observables Z y las no observables V.

- Previamente a realizar inferencia causal, se preguntará por el posible papel de las condiciones no observables V.
- Sabrá que la asignación aleatoria permite equilibrar a nivel poblacional las condiciones no observables V.
- Sabrá que la asignación aleatoria no garantiza un equilibrio absoluto entre las muestras en comparación.
- Recurrirá a los métodos de ajuste para equilibrar las condiciones Z relevantes.
- Recurrirá a la asignación al azar para equilibrar las condiciones V no observables y las Z poco relevantes.
- Definirá como transversal el estudio en que la recogida de los valores de X e Y sean simultáneos.
- Definirá como longitudinal el estudio en el que los valores de X e Y están separados por un lapso de tiempo suficiente para que los hipotéticos efectos de X actúen en Y.
- Definirá como prospectivo el estudio longitudinal en el que la recogida de la información sigue su orden natural (se observa X y se espera en el tiempo a que acontezca Y).
- Definirá como retrospectivo el estudio longitudinal en el que la recogida de la información invierte el orden natural (tras observar Y, se averigua X en el pasado).
- Distinguirá si el término control lo es de una causa (como en el ensayo clínico) o de un efecto (como en diseño de casos y controles).
- Se preocupará por la eficiencia de un diseño y de un análisis.
- Sabrá que los diseños apareados o los del estudio del cambio persiguen una mayor eficiencia eliminando la variabilidad interunidad.
- En un estudio que relacione dos variables se preocupará por recoger la información de cada una de ellas sin que la segunda variable pueda afectar los valores de la primera.
- Identificará como experimental el estudio en el que el investigador decida el valor de la causa en estudio (asigne el tratamiento).
- Percibirá las ventajas del diseño experimental en cuanto a equilibrar las variables desconocidas V y a optimizar la eficiencia de la estimación.
- Se preocupará porque el investigador que incluya los pacientes en el estudio desconozca la secuencia de aleatorización.
- Identificará los ensayos comunitarios como aquellos en los que la asignación del grupo de intervención se decide en conjunto para un grupo de unidades.
- Percibirá la grave amenaza de los desvíos del protocolo.
- Sabrá que sólo el análisis por intención de tratar (ITT) respeta el carácter experimental del estudio.

## Métodos de control de las terceras variables observables

Hay que tener muy claro qué papel desempeñan las diferentes variables en un estudio de intervención.

### Recuerde



*En este texto, se representa por Y a la variable respuesta; por X, a la intervención y por Z, a las condiciones de las unidades.*

### Ejemplo 12.1



Un tratamiento farmacológico o un consejo dietético pueden ser considerados como variables potencialmente intervinientes (X).

**Ejercicio 12.1**

Un estudio pretende evaluar el efecto que, sobre el nivel de cumplimiento de la medicación, tiene el número de visitas de seguimiento.

Suponga:

- a) un estudio experimental, que asigna a los pacientes un determinado número de visitas, y
- b) un estudio observacional, que simplemente observa el número de visitas espontáneamente solicitadas por el paciente.

¿El número de visitas es X o Z en cada estudio?

**Recuerde**

*Una variable susceptible de ser intervenida debe ser «asignable».*

El capítulo anterior se terminó definiendo el efecto causal ajustado por los valores de las condiciones Z. Antes de comentar las herramientas de las que dispone el investigador para dejar fijas estas variables Z, conviene recordar que un ajuste bien planificado suele conllevar un beneficio en la precisión de la estimación.

**Ejemplo**

En el ejemplo de la PAD, cabe esperar menos diferencias entre dos pacientes de 40 años tratados en el mismo centro que entre dos pacientes de diferentes edades y diferentes centros. La comparación directa de estos dos pacientes, al estar menos influida por fluctuaciones aleatorias, aportará más información.

**Recuerde**

*El ajuste permite controlar la variabilidad de la variable Z que se deja fija. Eso implica, en un buen diseño, mayor eficiencia estadística, por lo que un mismo número de casos permite un menor error de estimación y un intervalo de confianza más estrecho.*

Un ajuste no planificado en el protocolo puede tener consecuencias imprevisibles sobre la eficiencia; puede conllevar un beneficio o, por el contrario, aumentar considerablemente el error típico de estimación de los parámetros.

Asimismo, sólo el ajuste previamente planificado (y, por tanto, enmascarado con respecto a los resultados) permite garantizar la ausencia de sesgos (fig. 12-1).



**Figura 12-1** Ajuste por variables conocidas. Un profesional que desee tener controlada una variable conocida y medida sin error, utilizará un sistema que le garantice que queda igualmente distribuida en los grupos en comparación, de la misma forma que al freír un bistec nos aseguramos que cada lado se fría justo en la cantidad que deseamos.

### Recuerde



*Sólo el ajuste bien planificado puede garantizar dos beneficios: ausencia de sesgo y mayor eficiencia de las estimaciones.*

La tabla 12-1 resume las diferentes posibilidades para el ajuste distinguiendo entre aquellas planificadas en el momento del diseño y aquellas que se introducen durante el análisis. En los siguientes puntos se comentan sucesivamente.

### Restricción

La forma más sencilla de ajuste es la restricción de la población en estudio, bien sea mediante la definición de **criterios de inclusión** en el protocolo, bien sea mediante el **estudio de un subgrupo** en el momento del análisis.

### Ejemplo 12.2



Se limita el estudio a casos de una cierta edad y de un hospital concreto. Ahora, estas dos variables ya no pueden ser una explicación alternativa a la relación observada entre el tratamiento X y la respuesta Y.

Su gran ventaja es la sencillez de aplicación, mientras que su inconveniente es la reducción de la población objetivo, tanto para obtener casos para el estudio como para aplicar después sus resultados.

Opción	Fase	Nombre	Ventajas	Inconvenientes
<b>Restricción</b>	Diseño	Criterios de inclusión	Control completo Barato Simple de diseñar Simple de analizar	Reduce la población Número de variables limitado Posible confusión residual (si las restricciones son amplias)
	Análisis	Análisis de un subgrupo		
<b>Estudio por subgrupos</b>	Diseño	Bloques (apareamiento)	Potencia Eficiencia	Pierde flexibilidad Coste
	Análisis	Estratificación (apareamiento)	Sin premisas Directa Cálculo simple	Dispersión de casos en estratos Diferentes estratificaciones Difícil «sumarización»
<b>Modelado Estadístico</b>	Diseño	Modelado	Factible con pocos casos Redondea efectos menores	Muchas premisas Elección del modelo Elecciones variables Interpretación Parametrización del software
	Análisis	Covarianza, regresión, otros	Permite predicciones Permite variables continuas Permite varias variables	
<b>Ajuste global</b>	Diseño	Minimización	Permite considerar varias Z	Logística sofisticada
	Análisis	Pareja óptima	No reduce la población objetivo	

**Tabla 12-1** Opciones para el ajuste. Adaptada de Kleinbaum et al. (34)

### Ejemplo 12.3



En el ejemplo anterior, si un profesional de otro centro desea utilizar los resultados de ese estudio deberá analizar con detalle si existen características diferenciales entre ambos centros que puedan comprometer los efectos observados.

### Ejercicio 12.2



En un estudio experimental muy controlado, se ha logrado establecer el efecto de un tratamiento que disminuye el área necrosada tras un infarto. Los criterios de inclusión de dicho estudio exigían que el paciente fuera tratado en un plazo de tiempo muy limitado. Por esta razón, el protocolo del centro en el que usted trabaja no incluye este tratamiento para aquellos casos que superan este límite de tiempo. Usted desea tratar a este paciente. ¿Qué argumentos deben utilizarse?

**Definición**

Los **criterios de selección** son la aplicación de una restricción decidida en la fase de diseño del estudio.

**Lectura**

*La declaración CONSORT (9) recomienda no distinguir entre criterios de inclusión y exclusión, y llamarlos simplemente criterios de selección o de elegibilidad.*

Si una variable  $Z$  dispone de un amplio rango de variación y los criterios de inclusión son generosos, puede persistir una cierta confusión residual.

**Ejemplo 12.4**

Si la edad tiene un efecto claro en la respuesta, que hace muy diferente un caso con 18 años de uno con 70, limitar el estudio a pacientes de 18 a 70 es correcto, pero puede no ser suficiente.

**Definición**

El **estudio de un subgrupo** aplica la restricción en el momento del análisis.

Dada la ausencia de hipótesis previa, el análisis de un subgrupo debe siempre ser interpretado de forma exploratoria.

**Estudio de subgrupos**

La siguiente herramienta de ajuste es la definición de subgrupos, que reciben el nombre de **bloques** si se especificaron en el diseño, y **estratos** si se decidieron en el momento del análisis.

**Ejemplo 12.5**

En un ensayo clínico multicéntrico se especifica en el protocolo que se desea tener controlada la variable centro, y la asignación aleatoria se planifica de forma que en cada centro exista el mismo número de casos en cada grupo de tratamiento. Se dice que cada centro es un bloque y que está balanceado.

**Ejemplo 12.6**

En un estudio sobre el efecto de los programas de atención a domicilio en la supervivencia de las personas mayores que se quedan solas, se decide estratificar en el análisis por el nivel de actividades que realiza cada persona.

Si la variable que define los subgrupos no es modificadora del efecto, el principio de parsimonia aconseja proporcionar una **estimación única** y global del efecto que represente al efecto común de cada subgrupo.

**Recuerde**

*Sea parsimonioso al exponer sus resultados. Si es razonable, proporcione una única medida del efecto que sirva para todos los subgrupos, sean bloques o estratos.*

**Comentario**

*No hay nada más tedioso que un informe que va presentando resultados por subgrupos cuando el efecto es muy similar en todos ellos.*

**Nota técnica**

Una cuestión más delicada es ¿cuándo es razonable asumir que no hay interacción y que, por tanto, es lícito proporcionar una única medida? Desgraciadamente, la estadística es de poca ayuda en esta cuestión, ya que el tamaño muestral se define para el objetivo principal (p. ej., estudiar el efecto de una intervención), pero el estudio de la interacción (¿difiere este efecto por subgrupos?) requiere un mayor número de casos. Una prueba formal sólo detectaría interacción si ésta es de gran magnitud. Por ello, el consejo habitual del estadístico suele ser que las premisas del estudio conviene sustentarlas en datos previos, no sea que una sorpresa deje sin potencia el estudio.

**Ejemplo 12.7**

Si la presencia de alimentos en el tubo digestivo puede modificar el efecto del tratamiento, no tiene sentido combinar los resultados obtenidos en pacientes en ayunas con los de pacientes en plena digestión. Es más simple y directo definir al estado pandrial como criterio de inclusión. Si conviene, más adelante se estudiará la posible interacción o se complementará el estudio ya realizado con otro de la condición pandrial no considerada.

**Ejercicio 12.3**

Imagine dos estudios multicéntricos, ambos de  $n = 200$ , para comparar un tratamiento  $t$  con un control  $c$ . El estudio A ha sido diseñado para recoger información de cinco centros. Cada uno aporta 40 casos, 20 en cada brazo de tratamiento. El estudio B se inicia en veinte centros, de los que uno aporta 160 casos; y los otros 19 centros, los 40 restantes. Destacan 10 centros que sólo aportan un caso cada uno. ¿Cuáles de las siguientes afirmaciones son ciertas?

- El estudio A seguro que evita la posible confusión por la variable centro, ya que está equilibrado o balanceado.
- La razón entre el número de casos tratados con  $t$  y con  $c$  es constante para todos los centros del estudio A. Esto implica que la variable centro y la variable tratamiento son independientes (no son colineales), por lo que no puede haber confusión.
- En el estudio B, hay 10 centros que no pueden aportar, por sí solos, información del efecto del tratamiento y, por tanto, no podrán constituir cada uno de ellos un bloque separado.
- El estudio A, por su equilibrio en número de casos, parece que será más eficiente, proporcionando intervalos de confianza más estrechos.
- El estudio B al abarcar más centros, es más extrapolable a otros centros nuevos.
- El estudio B no es un estudio multicéntrico sólido.

El **apareamiento** es un caso extremo en el que cada bloque o estrato está únicamente compuesto por dos observaciones que comparten características que les diferencian de las otras parejas. El grado de conexión entre estas parejas, naturales o artificiales, puede variar.

**Ejemplo 12.8**

Los diseños con intercambio del tratamiento (*cross-over*) consisten en aplicar un tratamiento en un período y otro tratamiento en otro período, cambiando el orden de aplicación en diferentes pacientes. Cada paciente define una pareja de observaciones y constituye un bloque.

**Ejemplo 12.9**

En la búsqueda de causas  $X$ , a cada paciente que ha desarrollado cierta enfermedad ( $Y+$ ) se le busca una pareja sana ( $Y-$ ) que tenga los mismos valores en las terceras variables  $Z$ . Cada pareja constituye un estrato.

## Modelado

La siguiente opción de ajuste es el modelado estadístico, del que las diferentes técnicas de regresión son el procedimiento más habitual. El protocolo puede especificar en mayor o menor grado el modelo que se utilizará para realizar el ajuste.

### Ejemplo 12.10



Al comparar dos fármacos antihipertensivos (X), se puede especificar que el análisis de la respuesta (p. ej.,  $Y = \text{PAD}$  a las 12 semanas) se ajustará mediante un modelo de regresión lineal que incluye el valor inicial o basal (Z) de dicha presión.

Nótese que la relación entre la presión basal (Z) y final (Y) no es, ni mucho menos, el objetivo principal del estudio (un investigador experto en el tema debería conocer la magnitud de esta relación con anterioridad). El auténtico objetivo es conocer el efecto del tratamiento X en la respuesta Y ajustando por la condición Z.

Cuanto más **concreto** sea el protocolo sobre el modelo que se utilizará para realizar el ajuste, mayor será el carácter **confirmatorio** del estudio.

### Ejemplo 12.11



Un ejemplo muy sencillo de modelo que se especifica completamente en el protocolo es el llamado estudio del cambio o diferencia entre los valores finales y los iniciales o basales.

### Nota técnica



Este cambio, evolución o diferencia final-basal se corresponde (Senn, 67) con un modelo de regresión lineal en que el coeficiente («pendiente») de la relación entre la PAD a las 12 semanas y la PAD inicial vale 1.

Los paquetes de programas estadísticos dan una falsa sensación de facilidad en el uso de estos modelos de regresión. La selección de variables a incluir no es un tema trivial, especialmente en presencia de colinealidades o correlaciones entre las variables a introducir. Dejar el ajuste para el final es muy desaconsejable, ya que el margen de maniobra es siempre muy inferior.

**Recuerde**

*Si no quiere encontrarse con sorpresas al final del estudio, considere al inicio el método que usará para controlar las terceras variables.*

**Comentario**

*La práctica habitual de introducir en el modelo las variables que han resultado previamente significativas es muy peligrosa si se desea interpretar causalmente los resultados. Quizá podría ser correcta en un estudio exploratorio, que terminara proponiendo hipótesis que deberían ser ratificadas en estudios posteriores. Pero aun en esta situación, la colinealidad entre las variables introducidas generará problemas en la calidad de la estimación y en la interpretación de los resultados, como se comenta en el apartado «Papel del azar» (p. 302).*

**Ejercicio 12.4**

En el estudio del efecto de un tratamiento X en la PAD de las 12 semanas, se ha realizado un modelo de regresión que incluía la PAD en la semana inicial. Los estudios que se comentan a continuación han llegado al modelo por diferentes caminos.

Ordénelos de mayor a menor credibilidad que le ofrezcan las conclusiones.

Y remarque cuáles representan un salto importante en credibilidad.

- a) Tal y como especificaba el protocolo, la respuesta estudiada ha sido el cambio o diferencia entre la PAD final y la inicial.
- b) Con la PAD12 como respuesta, se ha introducido la PAD0 en el modelo con un coeficiente que ya venía especificado en el protocolo.
- c) Con la PAD12 como respuesta y tal como especificaba el protocolo, se ha introducido la PAD0 en el modelo y se ha dejado al programa que escoja el coeficiente por el procedimiento habitual (que minimiza el error de predicción de la respuesta).
- d) Como a pero sin que figurara en el protocolo.
- e) Como b pero sin que figurara en el protocolo.
- f) Como c pero sin que figurara en el protocolo.
- g) Tal como decía el protocolo, entre un montón de modelos posibles se ha seleccionado aquel en el que era más significativa la variable X.

### Ajuste global (minimización)

Finalmente, el cuarto método de ajuste considera simultáneamente varias variables en el ajuste. En un diseño experimental con aleatorización, persigue mejorar el equilibrio proporcionado por esta asignación al azar.

#### Ejemplo 12.12



En el estudio de la PAD se quiere tener controladas la edad, el centro, el estrés y el sedentarismo. En el momento de introducir un nuevo paciente en el estudio, el investigador proporciona los datos de estas variables a un centro de aleatorización que, sin romper el enmascaramiento del investigador, asigna el nuevo caso al grupo que consigue un mayor equilibrio en estas variables.

La minimización permite tener equilibrados los grupos sin necesidad de recurrir a criterios de inclusión más restrictivos ni a la definición de muchos subgrupos, lo que podría comprometer el reclutamiento de casos para un estudio.

#### Nota técnica



Los algoritmos habituales de **minimización** suelen ponderar por separado las variables consideradas en el proceso y persiguen un equilibrio en cada una de ellas.

La minimización puede combinarse también con la asignación aleatoria, manteniendo el enmascaramiento de los investigadores y de los pacientes.

#### Ejemplo 12.13



Un investigador interesado en ajustar por edad, género, estrés y sedentarismo (estas dos últimas medidas en escalas de 0 a 10) ha determinado en estudios previos que la relación de estas 4 variables con la PAD es la siguiente: cada año supone 1 mmHg más, los hombres tienen 8 mmHg más; cada punto de sedentarismo son 2 mmHg más y cada punto de estrés son 3 mmHg más. Así, un índice global combina estos valores en cada paciente y el próximo se asigna al grupo que hace más similares las medias de este índice en los dos grupos. Es óptimo, no para cada variable *Z* por separado, sino para el conjunto de todas ellas. Es decir, pretende compensar los desequilibrios de unas variables con los de otras.

También puede usarse en los estudios observacionales.

**Ejemplo 12.14**

En la misma búsqueda anterior de causas  $X$ , y de la misma forma que antes, a cada paciente que ha desarrollado cierta enfermedad ( $Y+$ ) se le busca una pareja sana ( $Y-$ ). La diferencia ahora es que la pareja debe tener un valor próximo a los valores de las terceras variables  $Z$ , sea en cada una de ellas, como en el ejemplo 12.12 o en su combinación en un índice como en el ejemplo 12.13.

## Técnicas de ajuste frente a la confusión y a la modificación del efecto

Las variables  $Z$  se incluyen en el modelo porque son condiciones previas sobre las que se estudia el efecto de la hipotética causa. Pero, por cuestiones logísticas, no se pueden considerar todos los atributos. ¿Cuáles seleccionar? La respuesta es: aquellos que puedan tener sus efectos confundidos con el tratamiento o bien aquellos que lo modifiquen.

La experiencia y conocimiento del investigador deben permitirle saber, a priori, qué variables  $Z$  predicen la evolución  $Y$  (y podrían, por tanto, acabar teniendo sus efectos confundidos con la intervención  $X$ ) y qué variables  $Z$  interactúan con el tratamiento  $X$  (y, por tanto, modifican el efecto de  $X$  en  $Y$ ).

**Recuerde**

*Concentre la atención y los riesgos estadísticos en el auténtico objetivo del estudio: la relación entre  $X$  e  $Y$  dadas las condiciones  $Z$ . (Las relaciones de  $Z$  con  $X$  e  $Y$  son una premisa, no un objetivo.)*

**Ejercicio 12.5**

Sean  $Z$  el centro,  $X$  el tratamiento e  $Y$  la evolución, ¿qué confunde a qué? Es decir, ¿qué tablas de las siguientes (tabla 12-2) tiene interés estudiar y presentar?

Si una variable  $Z$  es **modificadora del efecto**, es decir, si existe interacción, es complicado explicar en primer lugar un efecto común del tratamiento que debe matizarse en segundo tiempo según el valor de la variable  $Z$ . Es más sencillo explicar cuál es el efecto del tratamiento en cada nivel de la variable  $Z$ .

Z+	Y+	Y-	Z-	Y+	Y-		Y+	Y-
X+	6	5	X+	180	6	X+	186	11
X-	45	150	X-	150	20	X-	195	170
OR = 4			OR = 4			OR = 14,74		
Y+	Z+	Z-	Y-	Z+	Z-		Z+	Z-
X+	6	180	X+	5	6	X+	11	186
X-	45	150	X-	150	20	X-	195	170
OR = 0,11			OR = 0,11			OR = 0,05		
X+	Z+	Z-	X-	Z+	Z-		Z+	Z-
Y+	6	180	Y+	45	150	Y+	51	330
Y-	5	6	Y-	150	20	Y-	155	26
OR = 0,04			OR = 0,04			OR = 0,03		

**Tabla 12-2** ¿Qué tabla tiene más sentido estudiar? ¿Qué información aporta cada tabla?

### Ejercicio 12.6



Suponga los siguientes resultados sobre el efecto del alcohol, modificado por el género, en las habilidades motoras. Aunque los números (imaginarios) que se presentan son idénticos, ¿de qué forma es más fácil interpretarlos?

- Dos vasos de vino reducen un 40% la rapidez de reflejos, siendo esta reducción un 25% superior en las mujeres y un 25% inferior en los hombres.
- Dos vasos de vino reducen la rapidez de reflejos un 50% en las mujeres y un 30% en los hombres.

Si Z es modificadora del efecto, es preciso estimar más parámetros, lo que dificulta la interpretación y requiere más casos. Si una de las categorías de Z es especialmente relevante tiene sentido concentrar el estudio en esa categoría y dejar para más adelante el resto.

### Ejemplo 12.15



Se sospecha que cierto fármaco, que se usa muy poco, puede modificar el efecto de nuestro producto X en estudio. Será más práctico poner como criterio de exclusión la toma de ese fármaco y dejar para más adelante el estudio de la posible interacción.

## Interpretación de resultados ajustados

Se acaba de proponer el ajuste para estimar el efecto causal de una variable  $X$  habiendo eliminado el posible efecto de terceras variables  $Z$ . Veamos ahora cómo se deben interpretar estos resultados ajustados si la **variable  $Z$  es posterior** a la causa  $X$ .

### Ejercicio 12.7



Se sospecha que cierto gen ( $X_1 = \text{sí}$ ,  $X_2 = \text{no}$ ) provoca la aparición de enfermedades cardiocirculatorias ( $Y_1 = \text{sí}$ ,  $Y_2 = \text{no}$ ). Se sabe que las hiperlipemias ( $Z_1 = \text{sí}$ ,  $Z_2 = \text{no}$ ) originan también las enfermedades ( $Y$ ). Además, se sospecha que el gen  $X$  provoca las hiperlipemias  $Z$ . El orden natural de aparición de estas variables es  $X, Z, Y$ . Suponga que se ha decidido estudiar la relación entre el gen  $X$  y las enfermedades cardiocirculatorias  $Y$  a partir de los pacientes de un servicio ( $Z_1$ ) especializado en el tratamiento de pacientes hiperlipídicos. ¿A qué conclusión llegarían investigadores de este servicio que dispusieran únicamente de los datos de la subtabla  $Z_1$  (tabla 12-3)?

$Z_1$	$Y_1$	$Y_2$	$Z_2$	$Y_1$	$Y_2$		$Y_1$	$Y_2$
$X_1$	503	185	$X_1$	37	151	$X_1$	540	336
$X_2$	151	56	$X_2$	83	337	$X_2$	234	393
OR = 1,01 IC <sub>95%</sub> 0,7, 1,4			OR = 0,99 IC <sub>95%</sub> 0,6, 1,5			OR = 2,7 IC <sub>95%</sub> 2,2, 3,3		

**Tabla 12-3** Ejemplo 12-7, en el que la variable  $Z$  es posterior a la variable  $X$

### Ejercicio 12.8



Suponga que otros investigadores hacen el mismo estudio en la población general representada en la subtabla de la derecha. ¿A qué conclusión llegarían con estos datos?

### Ejercicio 12.9

Interprete el conjunto de los ejercicios 12.7 y 12.8.

### Recuerde



*Si las variables  $Z$  son posteriores a la causa  $X$  en estudio, la interpretación del ajuste por ellas es muy delicada.*

Conviene recordar que, en ocasiones, el objetivo es realizar una **predicción**, no una intervención.

**Ejemplo 12.16**

Recuperemos el ejercicio 11.3: la probabilidad de tener un hijo con el síndrome de Down es aproximadamente de 1 cada 600 embarazos en la población general, pero sube hasta alrededor de 1 cada 60 en las gestantes mayores de 40 años. Ahora bien, ¿y la edad del padre? La elevada correlación entre la edad de los padres puede originar que se estén confundiendo sus efectos y si, por ejemplo, se deseara conocer el efecto causal de la edad del padre, se debería ajustar por la edad de la madre. Ya se comentó en el ejercicio 11.10 que la única utilidad práctica (a nivel inmediato) de conocer el efecto de la edad del padre ajustado por la de la madre, sería poder decir a las futuras madres que se busquen un padre joven pero sin adelantar su propia edad de maternidad. En cambio, si la finalidad de establecer la relación entre la edad de los padres y el riesgo de síndrome de Down es poder seleccionar aquellos embarazos con elevada probabilidad de que el feto presente esta enfermedad, entonces el objetivo es predecir, y no es necesario buscar una relación causal por medio del ajuste.

Si el objetivo es realizar una **predicción**, se debe seleccionar, para hacer la predicción, aquella variable que sea más temprana en el tiempo, más fiable (en el sentido de obtener valores más similares en sucesivas determinaciones) e incluso más fácil de interpretar.

**Comentario**

*Si el objetivo **no** es intervenir, y desea seleccionar un conjunto pequeño de variables que permitan realizar una predicción o pronóstico, los métodos habituales de selección de variables disponibles en los programas de regresión multivariante pueden ser de gran ayuda para complementar los criterios clínicos.*

**Recuerde**

*Para realizar una predicción puede escoger aquella variable «más cómoda».*

**Ejercicio 12.10**

En el ejemplo del síndrome de Down, ¿qué ventajas puede tener la edad de la madre respecto a la del padre?

Volviendo al efecto causal, el atractivo del ajuste es que permite estudiar el efecto de la variable X, a la vez que se dejan igual, sin modificar, las variables por las que se ha ajustado. Esto es lo que pretende significar la expresión «el **efecto de X independientemente de Z**». Por ello, la decisión de ajustar o no por una tercera variable Z, que puede estar relacionada con el tratamiento X en estudio, depende de la capacidad futura de intervención: ¿se puede intervenir sobre X sin alterar simultáneamente Z? O por el contrario, ¿es más factible conseguir cambios simultáneamente en X y en Z que sólo en X?

### Comentario



*Recuerde el diseño experimental para distinguir el efecto de la edad de ambos progenitores: dos grupos de madres (jóvenes y adultas) se cruzan, al azar, con dos grupos de padres (jóvenes y adultos). Ahora, la edad de ambos progenitores es independiente, pero éste es un resultado experimental que no reproduce ninguna realidad concreta. La pregunta que contesta este diseño no tiene interés práctico en nuestro entorno cultural.*

### Ejemplo 12.17



Al estudiar el efecto del tabaco, se puede desear ajustar por el nivel de alcohol. Lo que será muy interesante si se dispone de una intervención sanitaria que consigue disminuir el hábito tabáquico sin alterar el hábito alcohólico. Pero si la intervención sanitaria consigue, al mismo coste, disminuir ambos hábitos tóxicos, tendrá más interés práctico conocer el efecto simultáneo de ambos.

Por supuesto, este principio de «caja negra» paga el precio de no satisfacer la curiosidad racional: «Al final de todo, ¿cuál es la auténtica causa de...?».

### Ejercicio 12.11



Suponga que los expertos afirman que dejar el tabaco a la vez que se modera el consumo de alcohol es prácticamente imposible y, por tanto, que no puede intervenir sobre ambas variables a la vez. ¿Tiene sentido estimar los efectos de una de estas variables a nivel fijo de la otra? (Es decir, ¿ajustará en su estimación?)

**Ejercicio 12.12**

Suponga que cierta teoría psicológica establece que es más fácil cambiar simultáneamente los hábitos sobre dieta y ejercicio que cambiar cada hábito por separado. ¿Conviene estudiar el efecto de la dieta a nivel fijo de ejercicio?

**Ejercicio 12.13**

¿Qué variables hay que tener controladas? Conteste cierto o falso a las siguientes afirmaciones.

- Hay que tener presente todo lo dicho a propósito del ajuste, de la confusión y de la modificación del efecto.
- No hay que ajustar por otras consecuencias  $Y'$  de la causa en estudio  $X$ .
- Sí que puede ajustarse por los atributos  $Z$  (características poseídas por los pacientes).
- Debe ajustarse por aquellos atributos  $Z$  que, si no quedan controlados, pueden comprometer, por su conocida relación con la respuesta, la inferencia causal.
- Se puede ajustar por otras concausas  $X'$  si se cree que la intervención sobre  $X$  podrá hacerse «independientemente» del valor de la concausa  $X'$ .
- Si es más fácil intervenir simultáneamente sobre  $X$  y sobre la con-causa  $X'$  que sobre  $X$  independientemente de  $X'$ , tiene más sentido no ajustar por  $X'$ .
- Si se puede intervenir sobre  $X$  sin alterar  $X'$  entonces puede tener sentido ajustar por  $X'$ .

**Ejercicio 12.14**

El servicio de infecciones de cierto centro dispone de un centenar de casos de sida ( $Y$ ) y desean averiguar el posible efecto del nivel de promiscuidad ( $X$ ) en la aparición de la enfermedad. Para ello, desean obtener unos controles sin sida ajustando por las siguientes características  $Z$ : preferencias sexuales, uso de drogas por vía parenteral, empleo de preservativo, edad y nivel socioeconómico. De todas ellas, la única de la que se sospecha que pueda ser modificadora del efecto de la promiscuidad es el uso del preservativo. El servicio de traumatología del mismo centro tiene recogidas, con el permiso de los pacientes, todas estas variables en un archivo informatizado. Diga, de cada frase, si le parece correcta o no.

- Tiene sentido hacer dos grupos según el uso del preservativo en los que se estimará, por separado, el efecto de la promiscuidad.

**Ejercicio 12.14 (Cont.)**

- b) Dentro de cada grupo de uso de preservativo, se busca un control de forma que sus diferencias sean mínimas en el resto de variables.
- c) Para ello, se recurre a la minimización: con la ayuda de un algoritmo informático, se forman las parejas de manera que las diferencias entre los valores de las restantes variables  $Z$  sea mínima.
- d) El criterio de apareamiento se ha realizado sin considerar (incluso, «a ciegas») el valor de la causa en estudio (el nivel de promiscuidad).

**Ejercicio adicional**

Busque o proponga un ejemplo de una intervención de salud para modificar hábitos de vida (tabaquismo, consumo de alcohol, dieta, sedentarismo, estrés, etc.). Defina el papel de cada variable ( $X$ ,  $Y$ ,  $Z$ ). Modifique el diseño de forma que cambie el papel de las variables en el modelo.

**Ejercicio 12.15**

Una discusión habitual es si conviene luchar contra la presión arterial sistólica o contra la diastólica. Unos trabajos incluyen una variable y otros a la otra. ¿Usted qué opina?

**Recuerde**

*Si el objetivo es intervenir mediante cambios en  $X$ , antes de ajustar por  $Z$  pregúntese si puede **modificar  $X$  a nivel fijo de  $Z$** .*

**Ejercicio adicional**

A partir de un EC real (o simulado), invente preguntas sobre los criterios de inclusión para que sus compañeros respondan si la razón para figurar como tales es por razones estadísticas (posibles confusoras o posibles modificadoras del efecto), éticas o logísticas.

### Las terceras variables no observables

El ajuste puede aplicarse sobre las variables  $Z$  que influyen en la respuesta siempre que su relación sea conocida y puedan ser medidas sin error. Sin embargo, en la respuesta pueden influir también otras variables no conocidas o con grandes dificultades en su medición. Llamémoslas  $V$ . Si estas variables  $V$  también estuvieran relacionadas con el tratamiento en estudio, podrían confundirse sus efectos.

#### Ejemplo 12.18



Un ejemplo podría ser el estrés, que aunque se sospecha su influencia, es muy difícil de medir y cuantificar.

Como estas variables  $V$ , por definición, no son observables, no se puede recurrir al ajuste o control. La única solución es asumir alguna **premisa**: alguna suposición sobre su existencia o sobre su comportamiento.

#### Lectura



Stone (68) menciona cuatro premisas (tabla 12-4) para poder inferir causalidad:

1. La primera y más simple es la **suficiencia del modelo**, que consiste en suponer que todas las variables importantes ya han sido consideradas en el ajuste. Y, por tanto, que las variables  $V$  no existen. Para ser verosímil, exige una gran predictibilidad de la respuesta, por lo que su uso está limitado a ciencias muy deterministas como, por ejemplo, la Física.
2. En Farmacología, y en aquellos entornos en los que es éticamente aceptable la comparación experimental entre los grupos, suele recurrirse a la **asignación aleatoria** de las causas en comparación. Esta randomización garantiza que ambos grupos vienen de la misma población y tienen, por tanto, idéntica distribución poblacional de todas las variables  $V$ . Se trata de la única «premisa» que un buen diseño puede establecer.
3. La tercera premisa, usual en Epidemiología, asume que las variables  $V$  están igualmente distribuidas o repartidas entre los grupos que se comparan ( $V$  y  $X$  son independientes, no hay colinealidad). Por tanto, los efectos de  $V$  y de  $X$  **no** están **confundidos**. Es decir: se asume que las variables que influyen en la respuesta o bien son observables ( $Z$ ) (y ajustaremos por ellas); o bien, si no son observables ( $V$ ), son independientes del tratamiento.
4. Finalmente, dado que puede darse aquella situación en la que la influencia conjunta en la respuesta de todas las variables  $V$  resulte equilibrada a pesar de que existan desequilibrios particulares en algunas de ellas, se ha propuesto basar la interpretación de los estudios observacionales en la premisa de **asignación ignorante** del tratamiento, que supone que la respuesta potencial, previamente a la administración del tratamiento, es la misma en ambos grupos.

Nombre	Premisa sobre las variables V	Entorno usual
Suficiencia del modelo	No hay variables V	Ciencias deterministas
Asignación aleatoria	Las V están equilibradas en los grupos	Farmacología
No confusión	Las que hay son independientes del tratamiento	Epidemiología
Asignación ignorante	Su influencia conjunta está equilibrada	Ciencias sociales

**Tabla 12-4** Premisas habituales sobre las variables no observables (Stone, 68)

#### Recuerde



Entre todas las condiciones que equilibran las variables V no observables destaca la **asignación aleatoria**, pues es la única que está en manos del investigador.

Un estudio que no controle las variables desconocidas V es *caduco* en el sentido de que un nuevo conocimiento científico que convierta una variable desconocida, V, en conocida, Z, obliga a repetirlo.

#### Ejercicio adicional



Proponga ejemplos basados en su experiencia para que sus compañeros discutan qué terceras variables podrían influir en los resultados y si éstas son observables o no. A su vez, deben indicar qué técnica de ajuste proponen para controlarla.

## Papel del azar

La inferencia estadística puede recurrir al azar en dos ocasiones, para seleccionar a las unidades o para asignar los tratamientos. Como se explica a continuación, en el primer caso se persigue representatividad, y en el segundo, comparabilidad (fig. 12-2).

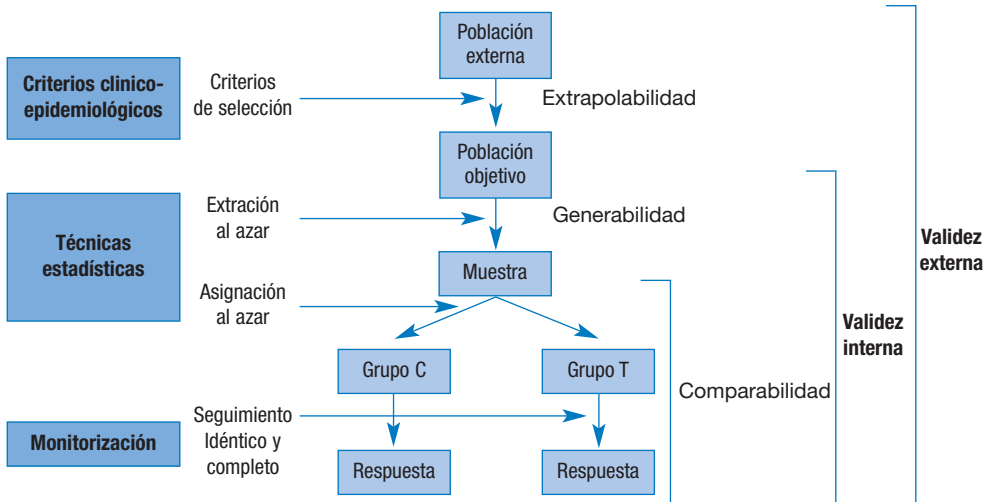
### Obtención al azar: representatividad

Ya se dijo que, si los elementos de la muestra eran seleccionados al azar, la teoría de probabilidad permite cuantificar (error estándar, intervalos de confianza) la incertidumbre al inferir al conjunto de toda la población los resultados obtenidos en la muestra.

#### Ejemplo 12.19



Al realizar la planificación sanitaria de una comunidad se recurre a una muestra aleatoria para conocer el estado de salud de sus ciudadanos.



**Figura 12-2** Dos aportaciones del azar a la validez: generabilidad y comparabilidad.

### Recuerde



*La teoría de probabilidad permite cuantificar las oscilaciones de los resultados debidas al azar del muestreo.*

### Ejercicio 12.16



Para realizar el estudio del ejemplo 12.19, a partir de un listado de todos los elementos de la población, se seleccionan al azar una serie de individuos en los que se evalúa su nivel de salud. Dado que es un derecho no participar, ¿qué consecuencias puede tener que algunos casos rehúsen?

### Recuerde



*La obtención al azar de la muestra persigue la representatividad o generabilidad de los resultados.*

### Asignación al azar (randomización): comparabilidad

Una vez se han seleccionado de la población aquellos individuos que formarán parte del estudio, el proceso de asignación aleatoria del tratamiento genera dos muestras de esta misma población. Dado que ambas muestras provienen, de manera aleatoria, de la misma población, la distribución de todas aquellas variables no intervenidas será la misma a nivel poblacional.

Dado que, a nivel poblacional, todas las variables están igualmente distribuidas en los dos grupos, no existe ninguna variable con la que se puedan confundir los efectos del tratamiento, por lo que la asignación aleatoria del tratamiento ofrece una estimación insesgada del efecto causal.

### Recuerde



*La asignación al azar del tratamiento persigue la comparabilidad de los grupos (fig. 12-3).*



**Figura 12-3** Papel del azar. Un profesional que desee tener *controlado* un elevado número de variables con pequeños efectos en la respuesta puede recurrir al azar para conseguir un reparto suficientemente similar del conjunto de todas ellas, de la misma forma que al freír unos champiñones introducimos entropía, azar, ruido, para que se hagan homogéneamente. ¡Y así resulta!

### Ejemplo 12.20



Imagínese una población con dos unidades: Montserrat Martí y José López (Rubin, 63). En el ejemplo de la PAD y los bloqueadores  $\beta$ , el efecto causal, en la unidad Montserrat Martí, de los bloqueadores  $\beta$  más dieta frente a la dieta sola, sería la PAD cuando se le asigna BB+D menos la PAD cuando se le aconseja D. En símbolos:

efecto causal en Montserrat Martí

$$e_{\text{Martí}} = \text{PAD}_{\text{BB+D, Martí}} - \text{PAD}_{\text{D, Martí}}$$

efecto causal en José López

$$e_{\text{López}} = \text{PAD}_{\text{BB+D, López}} - \text{PAD}_{\text{D, López}}$$

Y el efecto causal promedio en esta población de 2 unidades sería:

$$\begin{aligned} E[\text{PAD}_{\text{BB+D}} - \text{PAD}_{\text{D}}] &= \frac{1}{2} (e_{\text{Martí}} + e_{\text{López}}) = \\ &= \frac{1}{2} ([\text{PAD}_{\text{BB+D, Martí}} - \text{PAD}_{\text{D, Martí}}] + [\text{PAD}_{\text{BB+D, López}} - \text{PAD}_{\text{D, López}}]) \end{aligned}$$

**Ejemplo 12.20 (Cont.)**

Ahora bien, ya se vio que  $e_{\text{Martí}}$  y  $e_{\text{López}}$  no son observables a la vez y en las mismas condiciones. En un diseño experimental típico sólo se observaría o bien la muestra en la que se asigna el tratamiento a la unidad «Martí» y el control al individuo «López»:

$$\text{PAD}_{\text{BB}+\text{D}, \text{Martí}} - \text{PAD}_{\text{D}, \text{López}}$$

o bien la muestra con la situación contraria:

$$\text{PAD}_{\text{BB}+\text{D}, \text{López}} - \text{PAD}_{\text{D}, \text{Martí}}$$

Como estas dos muestras constituyen todas las posibles realizaciones muestrales, la esperanza del conjunto del experimento será:

$$\frac{1}{2} ([\text{PAD}_{\text{BB}+\text{D}, \text{Martí}} - \text{PAD}_{\text{D}, \text{López}}] + [\text{PAD}_{\text{BB}+\text{D}, \text{López}} - \text{PAD}_{\text{D}, \text{Martí}}])$$

lo que por simple reordenación de sus términos

$$= \frac{1}{2} ([\text{PAD}_{\text{BB}+\text{D}, \text{Martí}} - \text{PAD}_{\text{D}, \text{Martí}}] + [\text{PAD}_{\text{BB}+\text{D}, \text{López}} - \text{PAD}_{\text{D}, \text{López}}]) =$$

$$= \frac{1}{2} (e_{\text{Martí}} + e_{\text{López}}) = E[\text{PAD}_{\text{BB}+\text{D}} - \text{PAD}_{\text{D}}]$$

conduce al efecto causal promedio.

El conjunto de todas las posibles realizaciones muestrales se encuentra centrado alrededor del auténtico efecto causal, que es el valor que se desea conocer. Por ello, la asignación aleatoria permite una estimación insesgada.

Ahora bien, un buen investigador no tiene suficiente con el hecho de que la ciencia avance porque el conjunto de los estudios apuntan en la dirección correcta. Él quiere que su estudio concreto acierte en sus estimaciones (fig. 12-4).



**Figura 12-4** Límites del azar. Si una variable importante debe quedar equilibrada entre los grupos, el azar no lo garantiza, especialmente si la muestra es pequeña. El azar que le sirve al cocinero para *equilibrar* los champiñones, no funciona con los bistecs. Puede argumentarse que, si se dispusiera de muchos bistecs, la mitad de ellos se harían por su cara proximal y la otra mitad por la distal, llegando a un equilibrio a «largo plazo». Ciertamente, habría *equilibrio* para el conjunto de los comensales. Pero cada uno de ellos se comería un bistec que tendría una cara cruda y la otra hecha. De la misma manera, si un investigador quiere garantizar que en su estudio concreto una variable importante quede equilibrada, no debe descansar en el azar.

**Ejemplo**

En el **ejemplo 12.20**, un investigador concreto asignaría cada observación a un tratamiento, y en función del azar sólo observaría o bien el tratamiento en estudio en la primera unidad y el control en la segunda ( $PAD_{BB+D, Martí}$  y  $PAD_{D, López}$ ), o bien la situación contraria ( $PAD_{BB+D, López}$  y  $PAD_{D, Martí}$ ). En ambas posibilidades, el efecto del tratamiento estaría «confundido» con cualquier característica que diferencie los casos tratados de los controles. En este ejemplo, el género podría ser una explicación alternativa a posibles diferencias entre los casos.

**Recuerde**

*Aunque la asignación aleatoria permite estimaciones insesgadas a lo largo de todos los experimentos, **no garantiza** que cada posible experimento esté equilibrado para toda tercera variable.*

**Ejemplo 12.21**

(Prestado de S. Senn): A nivel global puede ser razonable aceptar que en 1 de cada 10.000.000.000 vuelos se averíen todos los motores, pero a los pasajeros de ese vuelo...

¿Qué puede hacerse? Muy sencillo: no dejar en manos del azar aquellas terceras variables que, si resultaran desequilibradas, pudieran comprometer las conclusiones del estudio. Nótese que las variables no observables  $V$  nunca comprometerán la credibilidad del estudio (porque no puede observarse su desequilibrio). Pero las variables  $Z$  sí que pueden observarse. Si una de ellas puede comprometer la credibilidad de los resultados, no conviene dejarla en manos del azar.

**Ejemplo 12.22**

Si el estadio inicial de la enfermedad es una variable crucial, que determina claramente la evolución de los pacientes, conviene controlarla por los métodos vistos en el ajuste de variables conocidas  $Z$ .

**Contraejemplo 12.23**

Si la situación personal, familiar y profesional es un conjunto de variables que influyen (un poco cada una de ellas) en la presión arterial, no conviene ajustar por todas ellas, ya que se haría inviable el estudio, al menos mientras no se defina una nueva variable que resuma todos sus efectos (fig. 12-5).



**Figura 12-5** Límites del ajuste. Pretender controlar cada pequeña variable haría inviable el estudio. Si se definen bloques de edad, género y color de los ojos en cada centro, siempre faltarán casos en alguno de los bloques y nunca se podrá cerrar el estudio.

### Recuerde



Las terceras variables *observables Z* se controlan con las técnicas de ajuste, y las *no observables V* con la asignación al azar.

### Ejercicio 12.17



De las siguientes expresiones, remarque aquellas que son ciertas.

- Si realizo al azar la asignación del tratamiento, las dos muestras, que vienen de la misma población, tienen una idéntica distribución poblacional.
- Si realizo al azar la asignación del tratamiento, las dos muestras, que vienen de la misma población, toman idénticos valores en ambas muestras.
- Si realizo al azar la asignación del tratamiento, la estimación es insesgada.
- Si realizo al azar la asignación del tratamiento, no debe preocuparme ningún posible desequilibrio entre las muestras, ya que la estimación es insesgada.
- Si mediante las técnicas de ajuste controlo todas las terceras variables *Z*, la asignación al azar ya no aporta nada al diseño.
- Si se acepta que la ciencia empieza por descubrir y conocer las variables más importantes, las técnicas de ajuste (que controlan las variables conocidas *Z*) tienen un papel más relevante que la asignación al azar (que controlan a las desconocidas *V*).

## Consideraciones éticas y papel del científico

Los principios éticos que rigen el diseño de los ensayos clínicos (autonomía, beneficencia y no maleficencia, equidad o justicia distributiva y confidencialidad), ¿cómo afectan al método de asignación al azar?

Nótese que de la «asignación al azar», el término que tiene implicaciones éticas es «**asignación**», pero **no** «**azar**». Si ambas pautas de tratamiento son éticamente asignables a los pacientes y ellos renuncian mediante el consentimiento informado a sus derechos de autonomía, ¿qué principio ético impide que sea el azar quien decida? Puede ser estéticamente feo, pero no es éticamente discutible.

### Recuerde



*«Asignar al azar» tiene implicaciones que pueden ser estéticamente feas, pero no éticamente reproables.*

### Ejemplo 12.24



El famoso ensayo clínico sobre la estreptomicina, realizado en los años cuarenta por el Medical Research Council (69), asignaba los pacientes al azar al grupo con antibiótico más reposo o bien al grupo tratado únicamente con reposo. Con los conocimientos de hoy en día, sobrecoge pensar qué procedimiento aleatorio pudo decidir sobre una evolución que implicaba la vida de esas personas. Dos consideraciones apoyan, una a la asignación, y la otra, a su aleatoriedad. La primera, común a casi todos los estudios, es que, con los conocimientos de aquel momento, los (posibles) efectos positivos se compensaban con los (posibles) efectos negativos y, en ausencia de conocimiento sólido, ambos tratamientos son éticamente asignables («equipoise»). La segunda consideración es un aspecto poco conocido de este ensayo: el Medical Research Council sólo disponía de estreptomicina para tratar a unos 150 enfermos, ante 300 candidatos. Y propusieron asignarlo al azar para garantizar el principio ético de equidad.

### Comentario



*El compromiso ético del ensayo clínico no viene impuesto tanto por la aleatoriedad como por la elección del tratamiento control (asignabilidad) y de las medicaciones de rescate.*

### Ejercicio 12.18



¿Es ético el uso de placebo?

## Lectura



*¿Hasta qué punto es lícito disminuir el procedimiento terapéutico habitual en un estudio? Por supuesto, ésta es una pregunta que debe ser contestada por los comités de ética y por los profesionales que consensuaron el protocolo que define dicho «procedimiento terapéutico habitual». En cualquier caso, conviene recordar que no sólo están en juego los derechos de los pacientes actuales con la enfermedad ya desarrollada, sino también los de los pacientes futuros. Por ello, suele decirse que, en situación de duda, la actitud más ética es empezar inmediatamente un estudio aleatorizado bien diseñado. Rawls (70) sugiere considerar la «situación de partida»: antes de saber el resultado del reparto de papeles que la vida nos depara, es decir, independientemente de si somos el paciente actual, el futuro o el investigador, ¿en qué sociedad preferimos vivir: en una que haga ensayos clínicos aleatorizados y aprenda rápidamente o en una que no los haga?*

Nuestra postura es que la actitud más ética consiste en empezar lo antes posible un estudio bien diseñado («asignar al azar el primer paciente») que permita que el máximo número de pacientes se beneficien de sus resultados (Cobo, 71).

## Recuerde



*Las obligaciones éticas de los investigadores y de los estadísticos son:*

- 1. Garantizar que el diseño permitirá contestar el objetivo del estudio, con el mínimo «coste» para los pacientes actuales.*
- 2. Iniciar, lo antes posible, dicho estudio.*
- 3. Ofrecer credibilidad en las conclusiones alcanzadas.*

## Criterios para clasificar un estudio

En este punto se definen dos claves para clasificar los estudios: la secuencia temporal y el plan de muestreo. A continuación se explican las implicaciones que tiene en el diseño considerar la eficiencia estadística y la calidad de la recogida de variables.

### Según la secuencia temporal

En el capítulo anterior se distinguió entre preguntas prospectivas (¿efecto de una causa?) y retrospectivas (¿causas de un efecto?). Ahora se verá que el diseño de recogida de datos puede también plantearse de forma prospectiva o retrospectiva. E incluso simultánea o transversal.

**Definición**

En un diseño **transversal**, la respuesta y la causa en estudio se miden en el mismo momento de tiempo.

**Ejemplo 12.25**

Se investiga en una serie de unidades, en el día de hoy, si tienen el sida (Y) y si son promiscuos (X). La limitación de este diseño proviene de que el sida de hoy puede ser una consecuencia de la promiscuidad de ayer, pero la (¿menor?) promiscuidad de hoy puede ser una consecuencia del sida de ayer.

En un estudio transversal, al realizar las mediciones en el mismo instante no se puede saber qué es lo que mueve qué: ¿qué hace de carro y qué hace de buey?

**Recuerde**

*Un estudio **transversal** no permite dilucidar el orden **causal**. En los estudios transversales debe poderse asumir que la causa X en estudio ha permanecido inmutable desde antes de que se desarrollara la consecuencia Y.*

**Comentario**

*Dado que esta invariabilidad sólo suele ser razonable para atributos inmutables, como el género, la repercusión científica de estos diseños es limitada.*

**Ejemplo 12.26**

Puede ser razonable asumir que el género de una persona ha permanecido invariable desde antes de desarrollar la enfermedad. Pero este carácter de invariabilidad es el que impide luego intervenir sobre el género («hágase usted mujer y disminuirá sus probabilidades de...»).

**Recuerde**

*La mayor limitación de los estudios transversales es que no contemplan el lapso de tiempo necesario para que la causa en estudio X ejerza su influencia.*

En un estudio longitudinal, la hipotética causa en estudio X está situada en el tiempo con anterioridad a la respuesta Y. Ahora bien, esta información puede recogerse en diferente orden.

**Definición**

En un diseño **retrospectivo**, tras observar la respuesta Y, se investiga en el pasado sobre los valores de la hipotética causa X.

**Ejemplo 12.27**

La pregunta sobre la promiscuidad (X) puede hacer referencia, no a la situación actual, sino a la situación en un pasado concreto, definido con un margen de tiempo suficiente.

**Definición**

En un diseño **prospectivo**, el registro de los datos sigue el orden natural de aparición de los acontecimientos: se observa primero la hipotética causa X y se espera en el tiempo a que acaezca la respuesta Y.

**Ejemplo 12.28**

Se selecciona a todos los pacientes actuales con sida de un cierto servicio que tengan normalizadas su cifras de CD4 y se les mide el grado de promiscuidad actual y el uso del preservativo. Al cabo de un cierto tiempo, se observa cuántos de ellos se han reinfectado por una nueva cepa del virus. El proceso del diagnóstico debe ser idéntico en ambos grupos y, a ser posible, el evaluador de la respuesta en estudio desconocerá los valores de promiscuidad y uso de preservativos.

**Ejercicio 12.19**

En el ejemplo anterior, no hay dudas de que la reinfección es la respuesta Y en estudio. Promiscuidad y preservativo, ¿qué papel desempeñan: son X o Z?

**Ejercicio 12.20**

En el hospital WW disponen de una lista de 100 pacientes que han padecido infarto. A cada uno de ellos se le busca un control de su misma edad, centro, obesidad y hábitos dietéticos y tóxicos. Toman, en todos ellos, una muestra de sangre y determinan la concentración del metabolito X. La comparación mediante la *t* de Student de datos apareados resulta significativa y concluyen que este metabolito es un factor que provoca el infarto. ¿Qué opina de esta conclusión? (Pista: piense en términos de causa (X) y efecto (Y) y en la secuencia temporal entre ambos): ¿cuándo se obtiene el metabolito X?

Un ensayo clínico es un diseño prospectivo, ya que se asigna la causa en estudio (X) y se espera en el tiempo para observar la respuesta (Y).

**Recuerde**

*Los ensayos clínicos son estudios prospectivos.*

Además, el ensayo clínico es un estudio confirmatorio, ya que parte de una hipótesis previa especificada en el protocolo.

**Recuerde**

*La especificación de la hipótesis en el protocolo permite el carácter confirmatorio de los ensayos clínicos.*

**Ejercicio 12.21**

Se han realizado 4 ensayos clínicos idénticos para valorar el efecto de un fármaco en el descenso de la presión arterial, habiendo resultado todos ellos estadísticamente significativos en esta respuesta, pero no en el descenso de los accidentes vasculares. Se sugiere que la potencia extra que proporcionaría un análisis conjunto de los cuatro estudios permitiría alcanzar la significación estadística. ¿Se trata de un estudio prospectivo o retrospectivo? ¿Es confirmatorio o exploratorio?

**Comentario**

Ciertos estudios **retrospectivos** utilizan bases de datos elaboradas de forma **prospectiva**. Tienen la ventaja de que las causas X están más libres de sesgos en su recolección, pero no garantizan que estén recogidas todas las variables relevantes. Además, para poder ser confirmatorios, deben especificar la hipótesis en el protocolo y justificar que ésta no ha venido sugerida por los resultados observados en esos pacientes, ya que entonces, la misma información habría sido utilizada para sugerir y confirmar la hipótesis, lo que se conoce como «capitalización del azar».

**Lectura**

Los términos “prospectivo” y “retrolectivo” suelen usarse también para distinguir entre estudios planificados antes o después de la ocurrencia de los datos. Pero este uso introduce ambigüedad, ya que estos términos suelen usarse para distinguir la secuencia temporal de la recogida de datos [...]. Para eliminar esta ambigüedad, los términos “prolectivo” y “retrolectivo” pueden emplearse y así evitar expresiones como [...] “estudio prospectivo retrospectivo”, [...] “prospectivo histórico”, “cohorte histórica” o “cohorte retrospectiva” para lo que aquí es denominado “cohorte retrolectiva” (Feinstein, 59).

**Según el plan de muestreo**

La recogida de información sobre las variables Y, X, y Z puede realizarse de varias formas.

**Definición**

En un estudio de cohortes, se recoge, en una población determinada, información en todas las unidades sobre las variables Y, X y Z.

**Ejemplo 12.29**

A partir de un censo de prostitutas masculinos de una cierta comunidad, se recoge información sobre sus hábitos y, en su momento, se valora su respuesta inmunitaria a los virus del sida y de la hepatitis.

Dado que ninguna de estas variables se ha fijado en el diseño, todas ellas ofrecen información sobre su distribución poblacional.

**Definición**

En un ensayo clínico se deja fija en el diseño la intervención X.

**Ejemplo 12.30**

Para estudiar el efecto de usar dos preservativos, se aconseja a la mitad de los prostitutas masculinos emplear uno; y, a la otra mitad, emplear dos.

**Ejercicio 12.22**

En este estudio, ¿informa la proporción de prostitutas que emplean dos preservativos de la correspondiente proporción poblacional?

**Definición**

En un diseño de caso-control se deja fija en el diseño a la respuesta Y.

De la misma forma que un ensayo clínico prospectivo necesita una muestra con la intervención control, un diseño retrospectivo, que parte de una serie de casos con la enfermedad (Y+), necesita una serie de controles sin la enfermedad (Y-).

**Nota técnica**

En un estudio prospectivo, se requiere un control para la intervención X, mientras que en uno retrospectivo se requiere el control para la evolución Y.

**Ejemplo 12.31**

No basta con observar que los pacientes con sida tienen un cierto grado de promiscuidad. Es necesario constatar que los que no tienen sida tienen menor promiscuidad.

**Recuerde**

*En un ensayo clínico, el término «control» indica la opción terapéutica o causa de referencia; pero en un estudio retrospectivo de casos y controles indica la ausencia de enfermedades o del efecto en estudio.*

**Ejemplo 12.32**

Se dispone de una muestra de 20 prostitutas que han desarrollado el sida (casos). Se selecciona otra muestra de 20 prostitutas que no lo han desarrollado (controles). En ambas muestras se les pregunta si han usado preservativos.

**Ejercicio 12.23**

Diga, de las siguientes proporciones, ¿cuáles pueden ser estimadas en el estudio del ejemplo 12.32 y cuáles no?

- La proporción de prostitutas que desarrollan el sida.
- La proporción de prostitutas con sida que usan preservativos.
- La proporción de prostitutas sin sida que usan preservativos.
- Entre los que usan preservativos, la proporción que desarrollan el sida.

En un diseño que fija alguna variable no pueden estimarse todos los parámetros.

**Comentario**

*Esta limitación influye en el análisis estadístico, ya que ciertos cálculos habituales, como por ejemplo los riesgos relativo y atribuible, no podrán hacerse.*

La *odds ratio* tiene la ventaja de que puede ser usada en cualquier tipo de diseño.

**Recuerde**

*En los diseños en los que se deja una **variable fija** hay algunas estimaciones que no pueden realizarse. La **odds ratio** no tiene esta limitación.*

En los estudios de casos y controles deben seleccionarse controles que sean «comparables» con los casos en estudio.

**Comentario**

*Por comparables se debe entender que puedan proporcionar una información insesgada sobre el efecto del tratamiento, es decir, que tengan los mismos valores en aquellas terceras variables que puedan originar confusión. En otras palabras, se desea que la única diferencia entre casos y controles sea, además de la presencia o ausencia de la enfermedad que los define, su grado de exposición al factor. Así, para poder hacer inferencia causal, se controlan las condiciones observables Z mediante las herramientas disponibles, y se debe asumir, por ejemplo, que las variables V no observables están igualmente repartidas entre los grupos (no confusión).*

**Ejemplo 12.33**



Para cada caso con sida se busca un control sin sida que tenga la misma edad, hábitos tóxicos, antecedentes patológicos, etc.

La tabla 12-5 muestra diferentes estrategias para obtener los controles, así como sus ventajas e inconvenientes.

Origen del control	Ventaja	Inconveniente
Población (p. ej., censo)	Seguro que todos los controles pertenecen a la población	Mala colaboración
Hospital: historia clínica, registros de enfermedades, etc.	Fácilmente accesibles	Mala representación
Vecindario		Mala colaboración Representación no asegurada
Amigos, familiares	Buena participación, fácilmente localizables	Representación no asegurada

**Tabla 12-5** Tipos de controles

**Recuerde**



*Los estudios de cohortes y los de casos y controles difieren en la perspectiva temporal y en el plan de muestreo (tabla 12-6).*

	$\geq 2$ muestras (fijan una variable)	1 muestra
<b>Retrospectivos</b>	Caso-control	
<b>Prospectivos</b>	Ensayo clínico	Cohortes

**Tabla 12-6** Perspectiva temporal y plan de muestreo en los ejemplos prototípicos de ensayos clínicos y estudios de cohortes y caso-control

La tabla 12-7 resume las propiedades principales de los estudios de casos y controles y de los de cohortes.

	Ventajas	Inconvenientes
<b>Casos y controles</b>	Rápidos de realizar Permiten obtener más casos	Menor fiabilidad de la información pasada El análisis estadístico puede sofisticarse
<b>Cohortes</b>	Son más confirmatorios Mejor información sobre X Mejores opciones de diseño	Más lentos Más caros Pueden requerir una cohorte grande

**Tabla 12-7** Propiedades de los estudios de cohortes y de casos y controles

### Eficiencia estadística

En los estudios de cohortes podría ser que sólo unos pocos individuos desarrollaran la enfermedad en estudio. Esta situación, muy deseable desde un punto de vista humano y sanitario, es fatal para la eficiencia estadística, que viene determinada por el tamaño del grupo menor.

#### Nota técnica



Si, por ejemplo, se desea comparar medias, la **eficiencia** estadística para una N total fijada se hace máxima cuando el tamaño n de los dos grupos es el mismo ( $n = N/2$ ).

Para optimizar la eficiencia estadística de la comparación o para evitar el coste de recoger datos que no mejoran la información disponible se puede recurrir a diferentes estrategias estadísticas de muestreo. Así, en caso de enfermedades raras, puede recurrirse, por ejemplo, a estudiar sólo una pequeña proporción de las unidades que no desarrollan la enfermedad.

#### Ejemplo 12.34



Se dispone de muestras de sangre congeladas para todos los individuos de la cohorte. Se estudian todos los casos que desarrollan la enfermedad, pero de los que no, se analizan únicamente los de una pequeña proporción («**casos y controles anidados en una cohorte**»).

#### Comentario



*El uso de las técnicas de optimización permite toda una serie de diseños alternativos que mejoran el coste y el seguimiento: «cohortes de casos y controles», «casos y controles híbridos», cohortes y casos, etc.*

#### Recuerde



*Para una misma «n» total es más **eficiente** aquel diseño y análisis que conduce a un menor error típico de estimación.*

El error típico de estimación y los correspondientes intervalos de confianza dependen del tamaño muestral y de la magnitud de la varianza de la variable respuesta. Por tanto, una forma de obtener un error típico de menor magnitud para un mismo número de observaciones consiste en disminuir la magnitud de esta varianza, eliminando, por ejemplo, la variabilidad entre individuos.

### Recuerde



Un estudio **apareado** pretende mejorar la eficiencia eliminando la variabilidad entre unidades.

### Ejemplo 12.35



Dos antifúngicos tópicos son administrados simultáneamente en dos zonas afectadas del mismo individuo y de tamaño similar. La respuesta es la reducción de la zona afectada al cabo de un tiempo determinado. La comparación directa de zonas de un mismo individuo permite eliminar la variabilidad de la respuesta entre diferentes pacientes.

### Ejemplo 12.36



En un estudio sobre el nivel de cumplimiento de la medicación, los pacientes utilizan alternativamente durante dos períodos de 6 meses dos modernos dispensadores de la medicación que permiten cuantificar el número de tomas dentro de un margen de tiempo preespecificado.

Para eliminar la variabilidad interunidades puede recurrirse a estudiar el cambio desde una medida inicial o basal de la respuesta; si bien, al valorar dos veces esta variable, la variabilidad intraindividual (que incluye el error de medida) afecta en dos ocasiones.

### Ejemplo 12.37



En la comparación de los dos antifúngicos tópicos anteriores, sólo se dispone de una lesión por paciente, pero la respuesta es la disminución del tamaño de la zona afectada desde los valores iniciales. Dado que el tamaño se valora al inicio y al final, el error de medida se comete en estas 2 ocasiones, a diferencia de un estudio que sólo mida el valor final que tendrá sólo el error de medida (y la variabilidad intraindividual) debida a la determinación final.

**Recuerde**

Para que un **estudio del cambio** (o evolución desde el nivel basal) sea más **eficiente**, la variabilidad entre unidad que se elimina debe ser superior a la intraunidad que se introduce (Senn, 67).

**Ejercicio 12.24**

En estos estudios, ¿qué afirmaciones son correctas?

- a) El beneficio del estudio del cambio es mayor si todas las lesiones tienen la misma magnitud, ya que de esta manera se hace menor la variabilidad entre individuos.
- b) El estudio del cambio suprime la variabilidad debida a diferentes condiciones iniciales en la lesión.
- c) El diseño apareado, respecto al estudio del cambio, presenta la ventaja de que también suprime la variabilidad debida a diferencias entre individuos durante el período de seguimiento.
- d) En el diseño del cambio, a diferencia del diseño apareado, no es necesario controlar aquellas variables (hábitos higiénico-dietéticos y tóxicos, por ejemplo) que podrían introducir diferencias durante la evolución.
- e) El diseño apareado, respecto al estudio del cambio presenta la ventaja de que no introduce la variabilidad interunidad.

**Calidad de la información**

Toda imprecisión en la recogida de datos implica pérdida de información y, en consecuencia, limita la eficiencia estadística, aumentando el error de las estimaciones.

Pero, además, si el error que se comente está ligado a las variables en estudio (X e Y), la estimación de la relación entre ambas estará sesgada.

La calidad de la recogida de la información sobre la hipotética causa X es menor en los diseños retrospectivos. Si además viene afectada por la consecuencia Y, se estará produciendo un sesgo o diferencia sistemática entre los grupos en comparación.

**Ejemplo 12.38**

Podría ser que la capacidad de recordar los hábitos sexuales pasados dependa de tener actualmente el sida. Un paciente que tiene 5 parejas por año puede no acordarse de ningún *partenaire* si no ha desarrollado el sida; pero, en caso de contagio, se «acordará» muy bien de cada uno de ellos.

En los estudios **prospectivos**, la calidad de la recogida de información es mayor. Los datos sobre todas las variables se recogen en su momento de aparición. La causa X no tendrá sesgos al valorarse antes de la aparición de la consecuencia Y. Es esta última la que puede recoger algún sesgo, especialmente si se trata de variables «blandas», que deban ser evaluadas con alguna carga subjetiva.

### Ejemplo 12.39



Un ejemplo sería la evolución de la enfermedad, a criterio del investigador o del propio paciente, valorada en escala ordinal (curación, mejoría, estable, etc.).

### Contraejemplo



Un contraejemplo sería la mortalidad. (Aunque sería discutible si hubiera que atribuirle el origen.)

Para minimizar estos sesgos, en los ensayos clínicos puede recurrirse al enmascaramiento del tratamiento, de forma que se pueda administrar y evaluar a ciegas, sin que sea posible identificar los dos tratamientos en comparación.

### Comentario



Se suele distinguir entre estudios **a ciegas** si cuando, o el investigador, o el paciente, desconocen el tratamiento administrado, y **con doble cegado** cuando ni uno ni otro lo conocen. Si ello no es posible y la evaluación no está libre de subjetividades puede recurrirse a la evaluación a ciegas. Para permitir este enmascaramiento debe recurrirse a la administración de un placebo que tiene idéntica apariencia exterior que el vehículo del tratamiento en estudio. Si se están comparando dos productos que se administran por vías diferentes puede recurrirse al doble placebo o **doble simulador** (double dummy), que consiste en administrar a cada grupo su tratamiento más el placebo-simulador del otro. Conviene mantener el enmascaramiento hasta que finalice la limpieza de la base de datos, que se **blinda** para no permitir modificaciones posteriores.

### Recuerde



Diga «a ciegas», «enmascarado» o «cegado» y evite «ciego».

### Historieta



Recuerde que blind  $\neq$  blinded.  
¿Apoya la gran cantidad de estudios realizados con investigadores ciegos y evaluadores ciegos las tesis de Ernesto Sábato en su «Informe sobre ciegos»?

**Recuerde**

En todo estudio que relaciona dos variables debe evitar que el error de medida de una de ellas guarde relación con la otra variable.

El **enmascaramiento** también es deseable en los estudios observacionales.

**Comentario**

Parte del llamado efecto placebo podría indicar la evolución debida a la autosugestión del paciente y del clínico (además de los efectos de la regresión a la media y de los otros tratamientos recibidos).

En las intervenciones que implican cambios de hábitos por el paciente, es imposible que éste desconozca el grupo de tratamiento al que pertenece. Si debe hacer ejercicio o dieta, tiene que saberlo. Y también el profesional que le asesora durante el período del estudio. En estos casos, conviene que el evaluador desconozca el grupo de tratamiento.

**Ejercicio 12.25**

Imagine un estudio en el que un grupo recibe consejo dietético y el otro no, con evaluación a ciegas. ¿Cuáles de las siguientes expresiones son ciertas?

- El efecto directo de la intervención y los posibles efectos indirectos originados por el hecho de conocer el paciente la intervención que se le aplica están confundidos, en el sentido de que no pueden descomponerse: el estudio proporciona una estimación de la suma de ambos efectos.
- Desde el punto de vista de adquisición de conocimiento, sería interesante poder distinguir entre los efectos directos e indirectos del ejercicio.
- El sesgo originado por el hecho de conocer el paciente su tratamiento elimina la utilidad posterior de los resultados.
- Dado que un paciente siempre sabrá si está haciendo ejercicio o no, preocuparse por este posible sesgo es irrelevante a efectos prácticos.

**Lectura**

La tabla 12-8 reproduce la traducción realizada por el doctor Esteve Fernández del tercer borrador de la guía de publicación STROBE sobre estudios epidemiológicos (72). A finales de 2007 se publicará la versión internacional definitiva.

	Punto n.º	Estudios de cohortes	Estudios de casos de controles	Estudios transversales
<b>Título y resumen</b>	1	(a) Identifique el artículo como un estudio de cohortes en el título o en el resumen  (b) El resumen debería ser una síntesis estructurada y muy informativa del artículo, teniendo en cuenta los puntos desarrollados en la lista de comprobación que sigue	(a) Identifique el artículo como un estudio de casos y controles en el título o en el resumen	(a) Identifique el artículo como un estudio transversal en el título o en el resumen
<b>Introducción</b>				
Antecedentes/fundamentos	2	Explique los antecedentes científicos y los fundamentos del estudio		
Objetivos	3	Consigne los objetivos específicos y cualquier hipótesis preespecificada		
<b>Métodos</b>				
Diseño del estudio	4	Presente los elementos clave del diseño del estudio. Declare el objetivo general del estudio original, si el artículo es uno de varios de un estudio en curso		
Marco	5	Describa el marco, lugares y fechas, definiendo los períodos de recogida de datos		
Participantes	6	(a) Explique los criterios de inclusión y exclusión, la fuente y los métodos de selección de los participantes  (b) Describa los períodos y los métodos de seguimiento	(a) Para los casos y los controles por separado, explique los criterios de inclusión y exclusión, la fuente y los métodos de selección  (b) Dé criterios diagnósticos precisos de los casos y motivos para escoger a los controles	Explique los criterios de inclusión y exclusión, la fuente y los métodos de selección de los participantes
Variables de interés	7	Proporcione una lista y defina claramente todas las variables de interés, indicando cuáles son tratadas como medidas de efecto o de exposición y cuáles son potenciales predictoras, variables de confusión o modificadoras		
Medidas	8*	(a) Proporcione detalles de los métodos de determinación para cada variable de interés (b) Si procede, describa la comparabilidad de los procedimientos en los diferentes grupos		
Sesgos	9	Describa cualquier tipo de medida utilizada para afrontar las potenciales fuentes de sesgo		
Tamaño de la muestra	10	Describa la justificación del tamaño del estudio, incluidas tanto las razones prácticas como las estadísticas		
Métodos estadísticos	11	(a) Describa todos los métodos estadísticos, incluyendo aquellos para el control de la confusión (b) Describa cómo se rataron las pérdidas en el seguimiento y los datos faltantes o perdidos (c) Si procede, describa los métodos para los análisis de subgrupos y los análisis de sensibilidad	(b) Describa cómo se trataron el emparejamiento y los datos faltantes o perdidos	(b) Describa cómo se trataron los efectos del diseño y los datos faltantes o perdidos
Exposiciones cuantitativas	12	(a) Explique cómo se analizaron las exposiciones cuantitativas; por ejemplo, qué categorizaciones o agrupaciones se escogieron y por qué (b) Si está indicado, presente los resultados de los análisis conjuntos o continuos y de los análisis agrupados		
Financiación	13	Explique la fuente de financiación y el papel de los financiadores en el presente estudio y, si fuera necesario, del estudio original en el que se basa		

<b>Resultados</b>				
Participantes	14*	(a) Para cada fase de estudio, informe del número de individuos: potencialmente elegibles, examinados para ser elegidos (si se conoce), incluidos en el estudio, que han acabado el seguimiento y analizados (b) Exlique las razones de no participación en cada etapa del proceso (c) Se recomienda la inclusión de un diagrama de flujo (d) Proporcione las fechas que definen el período de reclutamiento (e) En estudios apareados proporcione el número de controles por caso		
Datos descriptivos	15*	(a) Describa las características de los participantes en el estudio (p. ej., datos demográficos, clínicos o sociales) e informe sobre las exposiciones y los potenciales confusores (b) Indique, para cada variable de interés del estudio, la exhaustividad de los datos (c) Proporcione la media, la duración total y las fechas del seguimiento		
Datos de resultados	16*	Proporcione el número de sucesos del evento o tasas temporales	Informe del número de casos y controles en cada categoría de exposición	Informe del número de sucesos del evento o de medidas resumen para cada grupo en comparación
Principales resultados	17	(a) Proporcione medidas de asociación directas («crudas») o sin ajustar, así como ajustadas por las variables de confusión y su precisión (p. ej., intervalos de confianza del 95%). Deje claro por qué variables de confusión se ajustó y la justificación para incluir esas variables de ajuste y otras no (b) Para comparaciones que usen categorías derivadas de variables cuantitativas, indique el rango de valores o el valor mediano para cada grupo (c) Transforme las medidas relativas en diferencias absolutas de riesgo para periodos de tiempo significativos que no se extiendan más allá del rango de los datos (d) Proporcione los resultados ajustados por las variables de confusión para poblaciones objetivo realistas		
Otros análisis	18	Informe de otros análisis realizados, por ejemplo, análisis de subgrupos y análisis de sensibilidad		
<b>Discusión</b>				
Hallazgos clave	19	Resuma los hallazgos más importantes en relación con las hipótesis del estudio		
Limitaciones	20	(a) Discuta las limitaciones del estudio, teniendo en cuenta las fuentes potenciales de sesgo o de falta de precisión, y los problemas que pudieran derivarse por la multiplicidad de análisis, exposiciones y resultados estudiados. Discuta tanto la dirección como la magnitud de los potenciales sesgos (b) La discusión de las limitaciones no puede ser un sustituto de análisis cuantitativos de sensibilidad		
Generalización	21	Discuta la generalización (validez externa) de los resultados del estudio		
Interpretación	22	Aporte una interpretación global cauta de los resultados en el contexto de la evidencia disponible y las limitaciones del estudio, con especial atención a las interpretaciones alternativas		

\*Proporcione esta información por separado para casos y controles en estos estudios y, si es aplicable, para los grupos expuestos y no expuestos en los estudios de cohortes y transversales.

Fuente: STROBE statement. Checklist of essential items Version 3 (Sept 2005) [citado 15 Nov 2005]. Disponible en: [www.strobe-statement.org/PDF/STROBE-Checklist-Versión 3.pdf](http://www.strobe-statement.org/PDF/STROBE-Checklist-Versión%203.pdf)

**Tabla 12-8** Declaración STROBE: Lista de comprobación de puntos esenciales (Fernández, 72)

**Ejercicio 12.26**

De las siguientes características del ensayo clínico, ¿cuáles pertenecen también a los estudios epidemiológicos según STROBE?

- Hipótesis previa y, por tanto, protocolo.
- Diagrama de flujo de participantes.
- Controlar la multiplicidad de análisis.
- Enmascaramiento.

## Estudios experimentales

### Definición de ensayo clínico

La característica fundamental de los estudios experimentales es que el investigador determina el grado de exposición a la causa en estudio. En una palabra: puede *asignar* el tratamiento. El diseño experimental aplicado en Medicina recibe el nombre de ensayo clínico.

#### Lectura



*«Ensayo clínico es el método experimental aplicado a establecer los efectos de un procedimiento terapéutico. Se entiende por procedimiento terapéutico cualquier maniobra, intervención o producto químico o biológico administrado con cualquier intención, preventiva o curativa» (RD, 6).*

**Ejemplo 12.40**

Ya se ha comentado el ejemplo de la estreptomicina en el tratamiento de la tuberculosis (69), en el que se comparó la Estreptomicina más reposo en  $n = 55$  pacientes frente al reposo en  $n = 52$ . La respuesta era la evaluación enmascarada por dos radiólogos y un clínico de una placa radiográfica realizada a los seis meses de tratamiento.

#### Lectura



*La ventaja más importante de asignar al azar el tratamiento es que equilibra las terceras variables  $V$  no observadas. (Los estudios observacionales deben recurrir a asumir alguna premisa sobre estas variables  $V$ .) Además, esta asignación permite optimizar la eficiencia estadística, lo que en general implica asignar igual número de casos a los grupos. Finalmente, permite optimizar el ajuste de las terceras variables, de forma que mejoren también la eficiencia estadística. En cambio, los estudios observacionales, al ajustar, pueden resultar en grupos desequilibrados, con pérdida de eficiencia.*

**Recuerde**

Los diseños experimentales: 1) permiten controlar mediante la asignación al azar a las terceras variables  $V$  no observables; 2) permiten optimizar la eficiencia, y 3) disponen de mayor margen de maniobra para controlar a las terceras variables  $Z$ .

Todo ello repercute en un diseño más válido y más eficiente.

El carácter insesgado de estos estudios se expone de forma muy didáctica mediante los dos términos siguientes.

**Definición**

El **principio de comparabilidad** establece que dos o más grupos son comparables si difieren tan sólo en aquello que se desea comparar.

**Definición**

El **principio de intercambiabilidad** (*exchangeability*) dice que un diseño es insesgado o válido si se hubiera obtenido la misma estimación en el caso de que los dos tratamientos se hubieran intercambiado.

La tabla 12-9 resume algunas estrategias de asignación del tratamiento.

Asignación	Elección del criterio de asignación de tratamientos
Según criterio	El médico decide el tratamiento asignado a cada paciente
Aleatorización simple	Cada paciente es asignado a un grupo con la misma probabilidad $\pi$
Aleatorización balanceada por bloques	Cada «n» casos, el número de pacientes en cada grupo se equilibra. Los bloques se definen por terceras variables (orden, centro, edad, etc.)
Aleatorización por minimización	La probabilidad de asignar al paciente a un grupo se define de forma que, a la larga, se haga mínima la diferencia entre los grupos
Aleatorización según severidad ( <i>cut-off</i> )	Se procede en función de los valores de severidad. Por ejemplo, los casos severos se comparan con un activo, y los leves, con placebo
Aleatorización y libre elección ( <i>comprehensive cohort</i> )	Se dispone de datos de pacientes que han aceptado la asignación aleatoria y de datos de pacientes que han elegido libremente el tratamiento
Apostar por el mejor ( <i>Play the winner</i> )	Se da mayor probabilidad de asignar un paciente al grupo con mejores resultados hasta el momento

**Tabla 12-9** Posibilidades de asignación del tratamiento (Deeks et al., 73)

Para garantizar que las muestras generadas por la aleatorización son las que definen realmente los grupos, se debe evitar que el investigador que incluye los casos en el estudio conozca a qué grupo irán asignados antes de decidir incluirlos, ya que podría originar «sesgo de selección».

### Ejemplo 12.41



Durante las hambrunas de inicios del siglo xx, un estudio pretendía evaluar el efecto en el peso de los escolares irlandeses de un vaso de leche extra en la merienda. Al finalizar el estudio pesaban más los que no tomaron leche, ya que se dejó en manos del maestro la selección de los alumnos que sí tomaban y éste lo administró a los alumnos de menor peso.

### Recuerde



Para evitar el **sesgo de selección**, el grupo de asignación se enmascara o camufla (concealment) al investigador que incluye los pacientes.

### Lectura



Las recomendaciones CONSORT (9) dedican los apartados 8 a 10 a la «aleatorización»: «8) **generación de secuencia aleatoria**: método que se usa para generar la secuencia de asignación aleatoria, incluyendo detalles de cualquier restricción (p. ej., bloqueo, estratificación, etc.); 9) **ocultamiento de la secuencia aleatoria**: método que se usa para implementar la generación de secuencia aleatoria (p. ej., contenedores numerados o teléfono central), aclarando si la secuencia se oculta hasta que las intervenciones se han asignado; y 10) **implementación de la secuencia aleatoria**, quién genera la secuencia de asignación, quién selecciona a los participantes y quién asigna los participantes a cada grupo».

### Ejercicio 12.27



Compruebe si el trabajo de Chiasson et al. (74) cumple las directrices CONSORT sobre el informe de los métodos de aleatorización.

«De manera aleatoria, se asignó a pacientes idóneos, placebo o 100 mg de acarbosa 3 veces al día, tomada al principio de las comidas. Usamos un programa de ordenador para generar la secuencia de asignación aleatoria, estratificada por centro. La asignación aleatoria se llevó a cabo en bloques de 4 o 6. Se usaron envases de fármacos numerados para llevar a cabo la asignación aleatoria. Los pacientes se aleatorizaron secuencialmente en cada centro, pues el código de aleatorización se estratificó por centro. Las claves de aleatorización se ocultaron en un

**Ejercicio 12.27 (Cont.)**

envase etiquetado de tres compartimentos que se separó de los envases y se guardó en previsión de que los investigadores necesiten conocer el estatus de aleatorización del paciente. Un estadístico independiente, miembro del comité de revisión de la seguridad y calidad de datos, generó la secuencia de asignación. El reclutamiento y la aleatorización fueron gestionados en los centros.»

**Tipos de ensayos clínicos**

Se pueden realizar diferentes clasificaciones de los ensayos clínicos. La mayoría escapan de los objetivos de este curso.

**Comentario**

*Una clasificación muy utilizada tiene en cuenta la fase de desarrollo de un fármaco. Una vez se dispone de un producto químico o biológico ya «inventado», el desarrollo empieza con el estudio de los niveles tolerados en el organismo (fase I), prosigue con el estudio exploratorio (fase II) y confirmatorio (fase III) de eficacia y termina con el estudio de efectividad (fase IV). Todos aportan información sobre la tolerabilidad. La demostración de eficacia de la fase III pretende conseguir el registro del producto en el mercado (tabla 12-10).*

**Lectura**

*Lo que tiene **fase** es el **desarrollo** de un fármaco, no los ensayos clínicos que se utilizan en cada fase del desarrollo (ICH-E8, 75). Por ejemplo, en un fármaco ya comercializado puede desearse ampliar su ficha técnica con nuevas indicaciones o con nuevos estudios de tolerabilidad, o con estudios de interacción con otros fármacos, etc.*

Fase	Objetivo	Voluntarios	Tamaño	Diseño	Duración
I	Tolerabilidad y cinética	Sanos	Unidades	Dosis crecientes	Semanas
II	Exploración de eficacia	Pacientes	Decenas	Placebo, Aleatorizado, enmascarado	Meses
III	Confirmación de eficacia	Pacientes	Centenas	Placebo/Activo Aleatorizado, enmascarado	Años
IV	Efectividad y seguridad en grandes poblaciones	Pacientes	Millares	Activo	

**Tabla 12-10** Características de los ensayos clínicos según la fase de desarrollo del fármaco

## Comentario



Se suelen utilizar tres términos para determinar los efectos positivos, deseables del fármaco: la **eficacia** se establece en condiciones ideales (pacientes atendidos en un corto período de tiempo desde el inicio de los síntomas, con las mejores exploraciones complementarias...); la **efectividad**, en cambio, se estudia en condiciones reales, de práctica clínica habitual; finalmente, la **eficiencia** tiene en cuenta el coste de la intervención para la toma de decisiones. Su correcta evaluación se debe hacer considerando los incrementos de efectividad y coste respecto al procedimiento estándar actual, lo que se denomina **eficiencia diferencial**. Por otro lado, las técnicas **psicométricas** han permitido disponer de medidas validadas (insesgadas) y fiables (repetibles) de la **calidad de vida** de los pacientes, lo que facilita complementar la perspectiva clínica de eficacia con el punto de vista de los usuarios, beneficiarios últimos de las intervenciones sanitarias.

## Ensayo clínico comunitario

En el ensayo clínico habitual la asignación del tratamiento se realiza a los pacientes, que constituyen las mismas unidades en las que se obtendrá la evaluación de los resultados. Pero estas unidades podrían estar agrupadas en macrounidades en las que se asigna el tratamiento. Este tipo de diseños ha recibido los nombres de ensayos comunitarios o aleatorizados en grupo (*cluster*) o multinivel.

## Ejemplo 12.42



En un estudio sobre dos alternativas terapéuticas que precisan un elevado grado de entrenamiento de los profesionales, los centros sanitarios se asignan al azar y todos los pacientes de cada centro reciben el mismo tratamiento.

## Ejemplo 12.43



En un estudio de sensibilización de escolares al sida, se asignan los centros escolares a una u otra intervención y todos los escolares de un centro reciben la misma información.

## Ejemplo 12.44



En un estudio de prevención del tabaquismo, se desea comparar dos campañas publicitarias locales. Las comunidades se asignan al azar y luego se valora la respuesta en los individuos.

Nótese que en este tipo de diseños pueden establecerse mecanismos de «contagio» que rompan la independencia de las unidades y hagan que dos casos del mismo colectivo sean más similares entre sí que dos de diferentes colectivos, lo que debe considerarse en el análisis.

**Recuerde**

En un **ensayo comunitario**, la **asignación** del tratamiento se decide simultáneamente para un **grupo** de unidades.

**Ejercicio 12.28**

¿Son los estudios muticéntricos un caso particular de ensayos multinivel?

**Desvíos del plan del ensayo clínico**

Algunos casos no cumplen estrictamente el protocolo establecido (Cobo, 76). Esta situación, que sería incomprensible en un estudio sobre la calidad de fabricación de un producto, debe aceptarse en un ensayo clínico, dados los derechos de los pacientes incluidos en el mismo. Las desviaciones usuales consisten en pacientes que dejan de acudir a las visitas concertadas o no se toman la medicación prescrita.

**Recuerde**

Una buena intervención, un buen diseño, un buen investigador y un buen seguimiento conseguirán mantener al mínimo las desviaciones del protocolo.  
El crédito de los investigadores y la intervención se reduce a medida que aumentan los desvíos del protocolo.

**Comentario**

Algunos autores hablan de **quebrantamiento** (violation) del protocolo cuando el investigador incumple conscientemente el compromiso adquirido en el protocolo.

Las desviaciones del protocolo presentan dos grandes dificultades metodológicas. La primera es que pueden romper el equilibrio entre los grupos. La segunda, y definitivamente irremediable, es que impiden conocer el efecto real de la intervención. La potestad del clínico es aconsejar uno u otro tratamiento (X), pero el nivel de cumplimiento de las instrucciones pertenece al paciente (Z), y quizá sea una respuesta al tratamiento (Y'), pero no puede considerarse la causa X en estudio.

**Recuerde**

La variable experimental (X) es la asignación a un grupo de tratamiento. El cumplimiento del protocolo es una variable observacional (Z) que incluye atributos del paciente.

Por ello, el análisis principal de un ensayo clínico debe comparar a los pacientes a quienes se les ha recomendado un tratamiento con los que han recibido el otro consejo, lo que recibe el nombre de análisis por **intención de tratar** (ITT) o según la asignación del tratamiento (*as randomized*). El análisis alternativo, estudiando únicamente los pacientes que han cumplido el protocolo recibe el nombre de **análisis por protocolo** (PP). La medida del efecto futuro al asignar el tratamiento la proporciona el análisis por intención de tratar.

**Recuerde**



*Los desvíos del protocolo amenazan la validez que proporciona el diseño experimental y, por tanto, lo acercan al diseño observacional.*

**Comentario**



*Un farmacólogo puro puede argumentar que él desea conocer el efecto del tratamiento en aquellos casos que cumplieron el protocolo, pero un farmacólogo clínico se preocupará por el efecto en aquellos casos a los que se les asignó el tratamiento.*

**Ejemplo 12.45**



Se asignan al azar 200 casos a dos tratamientos alternativos A y B. En el grupo A, todos los pacientes cumplen y acaban el protocolo, mientras que en el grupo B, 50 de ellos dejan de tomar la medicación. En el grupo A se curan 60 y en el B, 40. (Tabla 12-11).  
Así, entre los que acaban (PP), el B tiene mayor eficacia, ya que son un 80% (40 de 50) respecto al 60% (60 de 100) del A. En cambio, respecto a los que empiezan (ITT), considerando que no tomar la medicación es un fracaso del tratamiento, es el grupo A el que tiene mayor eficacia, 60% respecto al 40% del grupo B.

	N	Pérdidas	Acaban	Curan	P(Cura   acaba)	P(cura   empieza)
A	100	0	100	60	60%	60%
B	100	50	50	40	80%	40%

**Tabla 12-11** El grupo B tiene un 50% de pérdidas y dificulta la interpretación de resultados

**Comentario**

*El análisis por cumplimiento del protocolo es más inmediato, pero no tiene interpretación posterior. En cambio, la gran dificultad del análisis por intención de tratar reside en que no se dispone de valores para los casos que no han cumplido el protocolo. Si hay que incluirlos en el análisis, ¿qué valor se les puede asignar? ¿Cómo tenerlos en cuenta? La solución más aceptada y popular consiste en copiar el último valor disponible (last observation carried forward o LOCF).*

**Ejemplo 12.46**

Si el paciente sólo participó en el estudio hasta el tercer mes, se asigna, a la evaluación de los 12 meses, el valor observado en el tercer mes.

**Comentario**

*El problema de los datos faltantes es muy serio y de imposible solución ya que, por definición, falta información que no puede «inventarse». Una elegante clasificación distingue entre datos ausentes completamente aleatorios (no dependen de nada); aleatorios (sólo dependen de otras variables observadas); o no ignorables (dependen de variables observadas y no observadas). Los primeros pueden generarse al azar, los segundos, en función de las variables observadas y los terceros son los únicos que no pueden generarse y pueden originar problemas. Pero la clasificación de un dato ausente en una de estas categorías debe hacerse por creencias del investigador, ya que no se dispone de las variables no observadas. Por tanto, esta clasificación descansa en premisas no contrastables, lo que le resta valor.*

Las conclusiones serán firmes si se pueden generar los datos ausentes bajo diferentes premisas y los resultados del análisis permanecen estables, independientemente de cómo fueron generados los datos. Ello requiere que el estudio tenga un número razonablemente reducido de pérdidas.

**Comentario**

*Dos argumentos aconsejan usar únicamente los casos que han cumplido el protocolo. El primero son los estudios de seguridad del fármaco, donde no tiene sentido incluir casos que, por ejemplo, no han tomado el producto, ya que seguro que no tienen acontecimientos adversos. El segundo argumento aparece en los estudios de equivalencia, donde añadir*

**Comentario (Cont.)**

*datos NO fiables tiende a diluir las diferencias entre los grupos. Incluir estos casos se convertiría en un premio para un mal diseño o un mal seguimiento, lo que de ninguna manera parece razonable.*

**Recuerde**

*La comparación más importante es **por intención de tratar**, si bien para el establecimiento de equivalencia o en el estudio de tolerabilidad, la comparación **por cumplimiento del protocolo** puede ser más relevante.*

*Ya se ha dicho que los desvíos implican pérdida de información. Cuanto menor sea, por tanto, mejor. El estudio sólo recuperará toda la credibilidad perdida si cualquier consideración sobre estos desvíos no conduce a conclusiones diferentes.*

**Recuerde**

*Conviene que ambos análisis (ITT y PP) lleven a la misma conclusión (Porta et al., 77).*

Son especialmente preocupantes las pérdidas relacionadas con la evolución, ya que podrían afectar de forma diferente a los grupos. Ello sucederá en los casos de eficacia nula o absoluta y de falta de tolerabilidad.

**Comentario**

*Si el paciente murió durante el período de seguimiento, por la sana costumbre de los médicos de no pasar visita a los muertos, no se tendrá un valor introducido en la respuesta al final del estudio; pero ello no implica que a ese paciente no le corresponda, en esa variable, un valor que indique fracaso.*

**Recuerde**

*Si la pérdida está relacionada con la evolución, se sabe mucho sobre el dato ausente y se debe intentar asignarle, a ciegas del tratamiento, un valor que considere esta información.*

Si las pérdidas son previas a la administración de un tratamiento enmascarado o se trata de pacientes erróneamente incluidos, podrían no afectar a la comparabilidad.

Finalmente, ¿qué se puede pensar de una situación en la que el paciente o no aparece o retira su consentimiento a participar en el estudio o simplemente no ha tomado toda la medicación? Todas estas situaciones pueden afectar a la comparabilidad.

### Recuerde



*Las pérdidas prealeatorización atentan contra la generabilidad de los resultados, y las pérdidas postaleatorización, contra la comparabilidad.*

### Comentario



*El uso de las tecnologías de la información ha permitido mejorar la calidad de los datos de forma muy notable. Quizá más por la posibilidad de establecer **filtros automáticos de coherencia** que permitan contrastar la información en el mismo momento en que se genera, que no por facilitar la transmisión o el almacenamiento.*

### Historieta



*Una anécdota del bioestadístico Dr. Alberto Cobos ilustra la importancia de estos filtros. El sistema avisó de que un paciente ya fallecido no podía tener presión arterial. Pero no era un error, estaba en espera de donación para trasplante. Al haberse ejecutado el filtro en el mismo momento en que se genera la información, la validación fue inmediata y no generó ninguna molestia.*

### Ejercicio 12.29



Para estudiar la protección para el cirujano del uso de un doble par de guantes (Cobo, 76), se ha realizado un ensayo (tabla 12-12) en el que los cirujanos son asignados al azar al grupo de utilizar uno o dos pares de guantes. Debido a ciertas creencias individuales (menor sensibilidad con dos pares, menor protección con un par, etc.), algunos cirujanos no han seguido el protocolo y han utilizado un número de pares diferente al asignado. En la tabla figuran los casos que han sido asignados a cada grupo, los pares de guantes que finalmente han usado y el número de casos en los que ha habido rotura y posible contaminación. Dadas estas transgresiones del protocolo se pueden plantear tres diferentes estrategias de análisis.

284 casos seleccionados				
Asignados a	Uno		Dos	
Casos	142		142	
Pares realmente usados	Uno	Dos	Uno	Dos
Casos totales	108	34	46	96
Se rompen	46	3	35	6
Resisten	62	31	11	90
% rotos	42,6%	8,8%	76,1%	6,2%

**Tabla 12-12** Datos del ensayo sobre el efecto de usar doble par de guantes

### Ejercicio 12.29 (Cont.)



1) Se incluyen todos los casos y se comparan los asignados a un par (49/142: 34,5%) frente a los asignados a dos pares (41/142: 28,9%), lo que se traduce en un  $IC_{95\%}$  de la diferencia de proporciones de  $-5,2$  a  $+16,4\%$  que incluye el 0 y es, por tanto, no significativo.

2) Se consideran evaluables sólo aquellos casos que han cumplido el protocolo y se comparan los asignados a uno que usan uno (46/108: 42,6%) con los asignados a dos que usan dos (6/96: 6,2%), lo que se traduce en un  $IC_{95\%}$  de 25,8 a 46,9%.

3) Se comparan todos los casos según los pares de guantes que realmente han usado: uno (81/154: 52,6%) o dos (9/130: 6,9%), lo que se traduce en un  $IC_{95\%}$  de 36,7 a 54,7%.

¿Cuáles de las siguientes son ciertas?

- El análisis por intención de tratar es el 1.
- El análisis por cumplimiento de protocolo es el 2.
- En el análisis 3 se está estudiando un atributo de los pacientes, no una variable experimental. Por lo tanto, la conclusión pierde la fuerza causal de que estaba dotada por ser ensayo clínico y debe ser considerado con la prudencia de los estudios no experimentales.
- En el análisis 1 se concluye que no se ha podido demostrar que aconsejar el uso de dos guantes mejore la seguridad de los cirujanos.
- En los análisis 2 y 3 los resultados son estadísticamente significativos.
- En el análisis 2, la selección realizada a los casos no garantiza que los grupos sean comparables. Se están eliminando la segunda y la tercera columna, que incluyen los casos más particulares. No es de extrañar que los «super precavidos» de la segunda columna (que aunque les toca usar uno, usan dos) tengan un valor muy bajo de rotu-

**Ejercicio 12.29 (Cont.)**

ras (8,8%); mientras que los «despreocupados» de la tercera columna (que deberían usar dos pero utilizan uno) tengan el porcentaje más alto de roturas (76,1%). Al eliminar estos casos tan diferentes, los resultados serán distintos.

g) En el análisis 3 se demuestra que si se aconseja en el futuro el uso de doble guante descenderá la proporción de roturas.

## Recapitulación: valor científico de los diferentes diseños

Ciertas características de los estudios científicos hacen que sus resultados sean más reproducibles.

### La asignación al azar es la base del proceso de estimación

**Ejercicio 12.30**

Una crítica habitual de los ensayos clínicos es que están realizados en unas condiciones alejadas de las reales de aplicación. Por ello, se sugiere realizar estudios más «pragmáticos» en los que evaluar también la eficiencia para permitir la posterior toma de decisiones. ¿Cuáles de las siguientes expresiones son ciertas?

- a) Algunos ensayos clínicos se hacen en condiciones que garantizan su validez interna (comparabilidad) pero que limitan su validez externa (representatividad o generabilidad).
- b) Puede mejorarse la validez externa sacrificando la validez interna.
- c) Para que la validez externa esté comprometida se requiere que las características diferenciales del ensayo sean una variable modificadora del efecto de la intervención.
- d) La asignación al azar es la principal responsable de la pérdida de validez externa, ya que las restricciones éticas que implica el azar limitan mucho la inclusión de casos.

**Recuerde**

*No se puede sacrificar la validez interna en nombre de la validez externa.*

Lectura



Se suele olvidar que el «error típico» de estimación de la diferencia de las medias de dos muestras se calcula, bajo la hipótesis nula, asumiendo que ambas vienen de la misma población, por lo que el error típico se limita a cuantificar la incertidumbre generada por el azar. Si no hay asignación al azar, ello implica que las dos muestras provienen de poblaciones que pueden diferir en algo más, lo que origina una incertidumbre adicional que no cuantifica el error típico ni el intervalo de confianza. Se ha demostrado que esta incertidumbre adicional genera, además de resultados sesgados, oscilaciones mayores de las cuantificadas por las medidas estadísticas habituales, fenómeno conocido como **sesgo impredecible** (73).

Recuerde



Eliminar la **asignación al azar** puede provocar estimaciones sesgadas tanto del efecto como de la variabilidad de su estimador.

Valor del diseño para imputar causalidad

Olkin (78) propone ordenar de menor a mayor según el nivel de evidencia científica para **imputar relación causal** los diferentes diseños según figura en la tabla 12-13.

Esta clasificación es útil, pero orientativa: conviene estudiar detalladamente cada estudio para poder valorar el nivel de credibilidad de su inferencia causal.

Comentario



Además de sus ventajas intrínsecas por asignación de la causa en estudio, los diseños experimentales suelen acompañarse también de: 1) un protocolo con una hipótesis que garantizan su carácter confirmatorio; 2) idéntico seguimiento y evaluación (enmascarada) de los grupos en com-

1	Informe de caso aislados
2	Series de casos sin controles
3	Series de casos con controles de la bibliografía o históricos
4	Análisis de bases informatizadas
5	Estudios retrospectivos de casos-controles
6	Series prospectivas
7	Ensayo clínico
8	EC confirmado incluyendo metaanálisis
9	Metaanálisis con datos originales

**Tabla 12-13** Jerarquía de los niveles de evidencia para establecer causalidad según Olkin (78)

## Comentario (Cont.)



paración; 3) un sistema de gestión de datos que mejora su validez y fiabilidad, a la vez que agiliza el proceso de la información; 4) una auditoría de calidad; 5) un registro que permite localizar los estudios para las evaluaciones globales; y 6) un permiso legal que regula la responsabilidad ética de los investigadores.

De la misma forma que los ensayos clínicos están beneficiándose de procedimientos de análisis clásicos en el entorno observacional, este último se beneficia cada día más de la metodología del entorno experimental.

## Ejercicio 12.31



¿Qué diferencia fundamental existe entre los métodos globales de epidemiología y el ensayo clínico controlado?

## Ejercicio 12.32

Explique brevemente la razón de ser del ensayo clínico controlado. Es decir: ¿con qué argumentos técnicos y éticos se justifica la experimentación en seres humanos?

## Ejercicio 12.33

Los grandes avances médicos en esperanza de vida están provocando un nuevo interés por la calidad de vida. Unos investigadores han desarrollado una escala que mide la calidad de vida de los ancianos con fractura de cadera. Han medido esta calidad de vida antes y después de la intervención quirúrgica mediante una  $t$  de Student apareada. Estiman, con una confianza del 95%, que la prótesis de cadera mejora la calidad de vida entre un 40 y un 50%. ¿Qué premisas son necesarias para poder afirmar que la cirugía es la causa de esta mejora?

## Recuerde



Al valorar la **calidad de un estudio** de inferencia causal conviene saber si:

- 1) es confirmatorio, con hipótesis previa expresada de forma unívoca;
- 2) es longitudinal y respeta el lapso de tiempo para que la causa actúe;
- 3) recoge sin sesgo las variables en estudio;
- 4) es experimental, ya que asigna la causa a las unidades;
- 5) controla mediante el ajuste las condiciones  $Z$  observables;
- 6) controla mediante el azar las condiciones  $V$  no observables;

**Recuerde (Cont.)**

- 7) tiene un diseño eficiente y un tamaño muestral adecuado que conduce a un error aleatorio reducido; y  
8) ha seguido con fidelidad el protocolo previsto.

**Lectura**

La tabla 12-14 (73) ordena los diseños para evaluar una intervención en función de la calidad o credibilidad de sus conclusiones.

**Ejercicio adicional**

Busque en internet la última versión, en inglés o en castellano, del Manual de la Cochrane Collaboration y conteste las siguientes preguntas:

¿Cuáles son los objetivos de esta Colaboración?

¿Qué apartados contiene el Manual?

¿Qué sesgos resalta en su capítulo sobre «Evaluación crítica de los estudios»?

Explique cada uno de estos sesgos a un compañero, poniendo ejemplos de situaciones en las que se podrían estar cometiendo.

**Diseños experimentales**

Un estudio en el que el investigador controla al menos algunas condiciones del propio estudio, especialmente las decisiones relativas a la asignación de los pacientes a los distintos grupos de intervención.

*1. Ensayo clínico con asignación aleatoria*

Los participantes son asignados aleatoriamente a los grupos de intervención o control, con seguimiento a lo largo del tiempo para determinar las posibles diferencias en los parámetros de evolución. El proceso de asignación aleatoria y el enmascaramiento de la asignación de los pacientes a los distintos grupos de intervención garantizan que, en promedio, los determinantes conocidos y desconocidos de la evolución se distribuyen de manera homogénea entre los distintos grupos.

*2. Ensayo clínico con asignación casi aleatoria*

Los participantes son asignados por el investigador a los grupos de intervención o de control, pero el método de asignación no es exactamente una asignación aleatoria genuina con enmascaramiento del proceso (p. ej., asignación en función de la fecha de nacimiento, el número de historia clínica hospitalaria, etc.)

*3. Ensayo clínico semiexperimental sin asignación aleatoria*

El investigador controla el proceso de asignación de los participantes a los distintos grupos, pero no lleva a cabo un procedimiento de asignación aleatoria (p. ej., la asignación se realiza en función de las preferencias del paciente o el médico). Difiere del «estudio de cohorte» por el hecho de que la intención es experimental más que de observación.

**Diseños de observación**

Estudio en el que se investiga la variación natural de las intervenciones (o la exposición) en los participantes en el estudio, con objeto de evaluar el efecto de las intervenciones (o la exposición) sobre distintos parámetros de salud.

*4. Estudio con determinaciones iniciales y finales*

Estudio de seguimiento sobre pacientes en los que se ha realizado una forma de intervención y sobre pacientes en los que no se ha efectuado dicha intervención, con cuantificación de la variable de evolución tanto en el momento de inicio del estudio como después del período de intervención, y con comparación de los cambios en las puntuaciones o, si los grupos son comparables, de los valores finales. También se puede considerar un diseño experimental si el investigador posee el control sobre la introducción de la intervención o bien puede realizar manipulaciones deliberadas en la misma.

*5. Estudio de cohorte concurrente*

Estudio de seguimiento en el que se compara la evolución entre los participantes que han recibido una intervención y los participantes que no la han recibido. Los participantes son estudiados durante el mismo período (de manera concurrente), bien de manera prospectiva o, con mayor frecuencia, retrospectiva.

*6. Estudio de cohorte histórica*

Variación sobre el estudio de cohorte tradicional en el que se establece el resultado de una nueva intervención aplicada sobre los participantes estudiados a lo largo de un período de tiempo concreto, y se compara dicho resultado con el obtenido en los pacientes que no recibieron la intervención en un período previo; es decir, los participantes no son estudiados de manera concurrente.

*7. Estudio de casos y controles*

Se identifican los pacientes con y sin una forma de evolución dada (casos y controles, respectivamente) y se comparan entre ambos grupos los resultados de la exposición a una intervención concreta.

*8. Estudio con control histórico previo*

Se realiza la comparación de los resultados obtenidos en los participantes del estudio antes y después de una intervención. Las mediciones previas y posteriores pueden realizarse en los mismos participantes o en muestras distintas. También se puede considerar un diseño experimental si el investigador tiene el control de la aplicación de la intervención o bien puede manipularla deliberadamente.

*9. Estudio con diseño transversal*

Evalúa en un momento concreto la relación existente entre la enfermedad y otras variables de interés existentes en una población definida.

*10. Serie de casos*

Descripción de un número de casos respecto a la forma de intervención y a los resultados (no se realiza la comparación con un grupo control).

Adaptada de CDR Report 4175.

**Tabla 12-14** Taxonomía de los diseños de los estudios para evaluar la efectividad de una intervención (Deeks, 73)

## Soluciones a los ejercicios

**12.1** En el estudio experimental, el número de visitas es la intervención  $X$  asignada a los pacientes. En el observacional, en cambio, se trata de un atributo o condición  $Z$  de los pacientes. Este atributo podemos utilizarlo para predecir el nivel posterior de cumplimiento (objetivo para el que usualmente empleamos las  $Z$ ). Pero si queremos que desempeñe el papel de una  $X$  y hacer inferencia causal, el tema es más delicado, ya que ahora es posible sospechar que el número de visitas es consecuencia de cierta característica personal que influye también sobre el nivel de cumplimiento de la medicación. Además, dado que este número de visitas no ha sido asignado, es preciso asumir que es posible hacerlo y que, cuando se haga, el paciente cumplirá con ellas. Nótese la importancia del término *asignación*. Incluso cuando no se acompaña del calificativo «aleatoria».

**12.2** ¡Atención! Esta pregunta tiene trampa (mis sinceras disculpas). Este capítulo discute un tema científico, de adquisición de conocimiento: ¿qué sabemos sobre un cierto tema? En cambio, la pregunta hace referencia a una decisión concreta: tratar o no tratar a ese paciente. Y la decisión requiere otra metodología. Por supuesto que el conocimiento sobre la eficacia influye en esta decisión, pero no es el único factor. A las consideraciones técnicas sobre hasta qué punto el lapso de tiempo extra que ha transcurrido puede comprometer esta eficacia, hay que añadir consideraciones sobre la seguridad (más fácilmente extrapolable) y, en especial, sobre el coste.

**12.3** Sólo es falsa la  $e$ ), ya que los resultados de este estudio se basan, en el fondo, en un único centro. Nótese que las respuestas  $a$ ) y  $b$ ) son dos formas correctas de expresar lo mismo.

**12.4** Ya están bien ordenadas.  $a$ ) y  $b$ ) ofrecen la misma credibilidad. Igual que  $d$ ) y  $e$ ). Los dos saltos más importantes son de  $c$ ) a  $d$ ) porque deja de estar especificado en el protocolo, y el último, ya que ese criterio es una auténtica barbaridad, algo así como si dijéramos: «No pares de analizar los datos hasta que den el resultado que yo quiero».

**12.5** Como el objetivo es estudiar la relación entre  $X$  e  $Y$  ajustando por  $Z$ , nos centraremos en la primera fila de tablas. La segunda fila de tablas aclara que  $X$  y  $Z$  están relacionadas: son colineales. Y la tercera fila que  $Z$  e  $Y$  están también relacionadas: la condición  $Z$  es predictora de la respuesta  $Y$ .

La estimación de la relación parece ser diferente al condicionar por centro ( $OR = 4$  en las dos primeras subtablas) que al considerar globalmente los casos ( $OR = 14,74$ ), por lo que tiene sentido condicionar por centro ya que es una característica previa de los casos: cada paciente entrará y saldrá del estudio condicionado a que es de su centro. Así, los resultados de interés son los de las dos primeras subtablas. Como el centro no es modificadora del efecto, se presenta un único valor del efecto ( $OR = 4$ ) mediante la combinación de las dos observaciones en ambos grupos de centro. (¡Y ojalá esto coincida con lo especificado en el protocolo!)

**12.6** Es más sencillo y legible el segundo sistema.

**12.7** En los pacientes del servicio de lípidos ( $Z_1$ ) no se observa relación entre la presencia del gen y el desarrollo de las enfermedades. Dado que el intervalo de confianza es razonablemente pequeño, se pueden descartar valores elevados de relación y no parece que una muestra mayor pudiera cambiar mucho las conclusiones.

**12.8** En la población general, estudiando conjuntamente los que tienen alterados los lípidos y los que no los tienen, se detecta una relación significativa entre el gen y las enfermedades cardiovasculares.

**12.9** Al condicionar por lípidos, no se observa la relación global establecida entre gen y evolución. Por lo tanto, parece que todo el posible efecto de los genes en la evolución viene mediado por los lípidos. Si se pudiera intervenir sobre los lípidos puede esperarse que se consiga anular el efecto de los genes. La moraleja de estos 3 ejercicios es que el ajuste por una variable posterior es mucho más delicado de interpretar.

**12.10** Si se acepta que el objetivo es identificar aquellos embarazos con una probabilidad de que el feto presente síndrome de Down lo suficientemente elevada para compensar los riesgos de exploraciones complementarias cruentas, se puede escoger para hacer la predicción, aquella variable que sea más barata, más temprana en el tiempo, más fiable (en el sentido de obtener valores más similares en sucesivas determinaciones) e, incluso, más fácil de interpretar. En todas estas propiedades, la edad de la madre antecede a la del padre.

**12.11** Sí. Dado que no puede intervenir a la vez sobre ambas, puede tener sentido estudiar sus efectos independientes de cara a priorizar las intervenciones.

**12.12** A nivel práctico, no. Quizá para satisfacer la curiosidad racional y contestar la pregunta retrospectiva sobre «¿cuáles son las causas de...?», pero no de cara a la intervención sanitaria.

**12.13** Todas son ciertas.

**12.14** Todas tienen sentido.

**12.15** Ahora el objetivo pretende ir más allá de la predicción, ya que se quiere modificar el riesgo vascular interviniendo sobre la presión arterial. La pregunta crucial es ¿se puede bajar una de estas presiones dejando fija la otra? Si las intervenciones disponibles van a actuar disminuyendo ambas, no tiene sentido estudiar el efecto de bajar una de ellas dejando fija la otra. Una posibilidad sería considerarlas simultáneamente, por ejemplo, mediante la presión arterial media.

**12.16** Evidentemente, comprometen la representatividad de la muestra. Convendría estudiar si tienen características diferenciales con los que sí han aceptado participar.

**12.17** Son falsas la *b*), la *d*) y la *e*). Para saber si es cierta la *f*) haría falta confirmar la hipótesis de que el conocimiento de las variables predictoras avanza de las más importantes a las menos importantes.

**12.18** Otra pregunta con trampa (mis disculpas). El placebo es tan sólo un simulador para enmascarar el tratamiento. La pregunta correcta es si el tratamiento que recibe el grupo control está de acuerdo con los protocolos profesionales. Recuérdese que al grupo «placebo» no se le suprimen los tratamientos de conocida eficacia. Por eso, muchos autores prefieren llamarlos estudios «*add-on*» para recordar que lo que se hace es añadir el tratamiento en estudio.

**12.19** Esta pregunta es más de ética que de estadística. ¿Cuál de las dos variables es más fácilmente intervenible sin atentar contra los derechos humanos que establece el principio de autonomía? Recuérdese la estrategia de los investigadores del asma y la soja en Barcelona: «Ya que no podemos cambiar los vientos ni interrumpir la descarga de soja, repararemos los silos y pondremos una lona para aislar esta descarga».

**12.20** Es una barbaridad. El metabolito, para poder provocar el infarto, debería ser anterior en el tiempo, no posterior. Para poder aceptar estas conclusiones es necesario asumir que el valor de este metabolito en este paciente era el mismo antes del infarto. Además, como el estudio no es aleatorizado, debe aceptarse que no hay confusión (es decir, que las variables no controladas por el diseño y el análisis están igualmente repartidas entre los grupos). Si se deciden a publicar estos resultados, se deberían discutir estas premisas para que los lectores puedan opinar sobre su verosimilitud.

**12.21** Aunque los datos han sido recogidos de forma prospectiva, la hipótesis no figuraba en los protocolos respectivos, por lo que se trata de un análisis exploratorio, que puede generar hipótesis («nuestros resultados sugieren...») pero no teorías («hemos demostrado que...»). Si la hipótesis hubiera figurado en los protocolos originales, debería haberse repartido el riesgo  $\alpha$  entre todos los contrastes de hipótesis que llevan a la misma decisión.

**12.22** Evidentemente que no, la variable  $X$  ha quedado fija por diseño. No es una variable aleatoria que estime nada.

**12.23** Desarrollar el sida es una variable fija, por lo que no se pueden estimar ni  $a$ ) ni  $d$ ), pero sí  $b$ ) y  $c$ ). Nótese que para ofrecer una estimación global de los prostitutos que usan preservativos deberían ponderarse las estimaciones  $b$ ) y  $c$ ) por la prevalencia de la enfermedad.

**12.24** Son falsas la  $a$ ), ya que si los individuos son muy homogéneos, el cambio eliminaría una variabilidad pequeña; y la  $d$ ), pues siempre es necesario que los grupos en comparación tengan el mismo seguimiento y evaluación.

**12.25** Todas figuran explícitamente, menos el enmascaramiento, que estaría implícito en medidas para evitar sesgos.

**12.26** La única falsa es la  $c$ ). El argumento  $d$ ) es correcto en el caso de un hábito. Nótese que, por el contrario, en el caso de un fármaco, es imprescindible poder distinguir ambos efectos.

**12.27** Los cumple.

**12.28** No, siempre que la asignación se haga a nivel de los pacientes.

**12.29** La única falsa es la última.

**12.30** *a)* y *c)* son correctas; *b)* es falsa, ya que para poder generalizar algo a un grupo más extenso, primero debe poderse aplicar al grupo más reducido. Y también es falsa la *d)*, ya que los principios éticos afectan al término asignación. Ya se dijo que una vez se ha decidido que una intervención es éticamente asignable a un paciente, añadir el calificativo «aleatoria» tiene connotaciones estéticas pero no éticas.

**12.31** La asignación del tratamiento.

**12.32** Ante la duda de asignar un nuevo tratamiento, la mejor forma de evaluar su efecto es precisamente asignándolo. Todos los tratamientos que se asignan a los pacientes deben estar de acuerdo con los protocolos estándar del entorno. Además, el ensayo debe estar bien diseñado para poder aplicar de la forma más inmediata sus conclusiones.

**12.33** Es necesario poder asumir que la calidad de vida de estos pacientes no hubiera evolucionado nada en ausencia de intervención. Esta premisa debe ser explícita en la discusión final del artículo. También es necesario justificar que no hay errores de medida ni de diagnóstico.