DECIMOSÉPTIMA EDICIÓN

ET

AMIR

MANUAL DE

ESTADÍSTICA y EPIDEMIOLOGÍA

Dirección editorial

EDUARDO FRANCO DÍEZ (5)
AIDA SUÁREZ BARRIENTOS (37)
MIKEL MAEZTU RADA (27)
PILAR PÉREZ GARCÍA (8)
JAIME CAMPOS PAVÓN (9)

MANUAL AMIR
Estadística y Epidemiología
(17.ª edición)

ISBN
978-84-19895-26-4

DEPÓSITO LEGAL
M-22209-2023

ACADEMIA DE ESTUDIOS MIR, S.L.
www.academiamir.com
info@academiamir.com

DISEÑO Y MAQUETACIÓN
Equipo de Diseño y Maquetación de AMIR.

Nuestra mayor gratitud a Hernán R. Hernández Durán, alumno AMIR, por haber realizado de manera desinteresada una revisión de erratas de nuestros manuales, que ha permitido mejorar esta 17.ª edición.

Este manual ha sido impreso con papel ecológico, sostenible y libre de cloro, y ha sido certificado según los estándares del FSC (*Forest Stewardship Council*) y del PEFC (*Programme for the Endorsement of Forest Certification*).

ET

Estadística y Epidemiología

Orientación MIR

[3,4]
Rendimiento por asignatura
(preguntas por página)

[14,6]
Número medio de preguntas
(de los últimos 11 años)

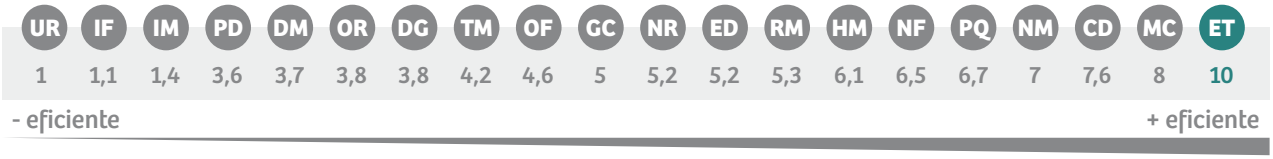
[10]
Eficiencia MIR
(rendimiento de la asignatura corregido por su dificultad en el MIR)

Estadística y Epidemiología es actualmente una asignatura de **importancia intermedia** dentro del examen MIR. Previamente era la segunda asignatura por detrás de Digestivo, pero en las últimas convocatorias el número de preguntas se ha reducido aproximadamente a la mitad (número de preguntas esperable: **entre 7 y 10**).

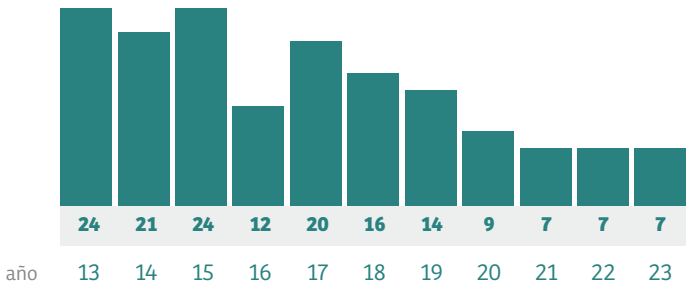
El tema estrella es el de **Tipos de Estudios Epidemiológicos**, que incluye preguntas teóricas. También son muy importantes temas en los que pueden caer problemas: **Medidas en Epidemiología** y **Estudio de una Prueba Diagnóstica**. Dentro del bloque de Estadística, lo más importante es Contraste de Hipótesis, si bien llama la atención que **no ha caído ninguna pregunta del bloque de Estadística en los últimos 4 años en el MIR**. Por otra parte, últimamente han aparecido en el MIR preguntas vinculadas a imágenes (interpretación de resultados y gráficos de estudios epidemiológicos).

La asignatura tiene una **alta rentabilidad** de estudio al ser la mayoría de conceptos repetidos y similares año tras año. El manual debe trabajarse de manera distinta al resto: cualquier tema de este manual se debe estudiar al detalle (salvo el tema 4), y por ello el manual está estructurado con **dos colores de texto**: texto en negro, que se debe estudiar íntegro; y texto en color aguamarina, que es simplemente texto aclaratorio o con ejemplos, que no se debe estudiar.

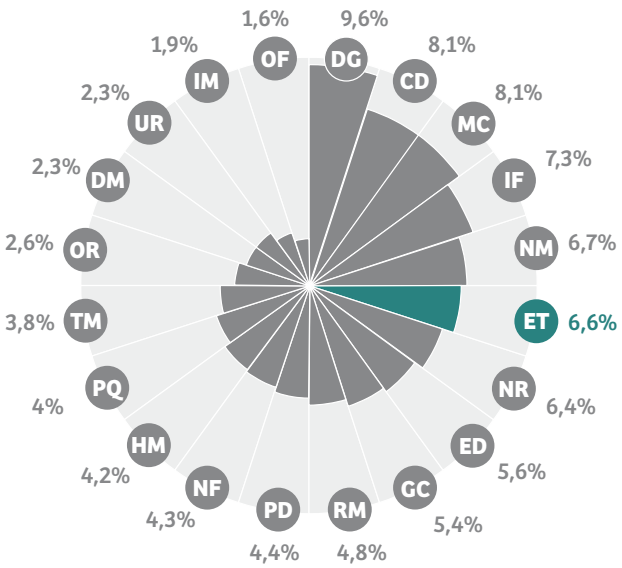
Eficiencia MIR de la asignatura



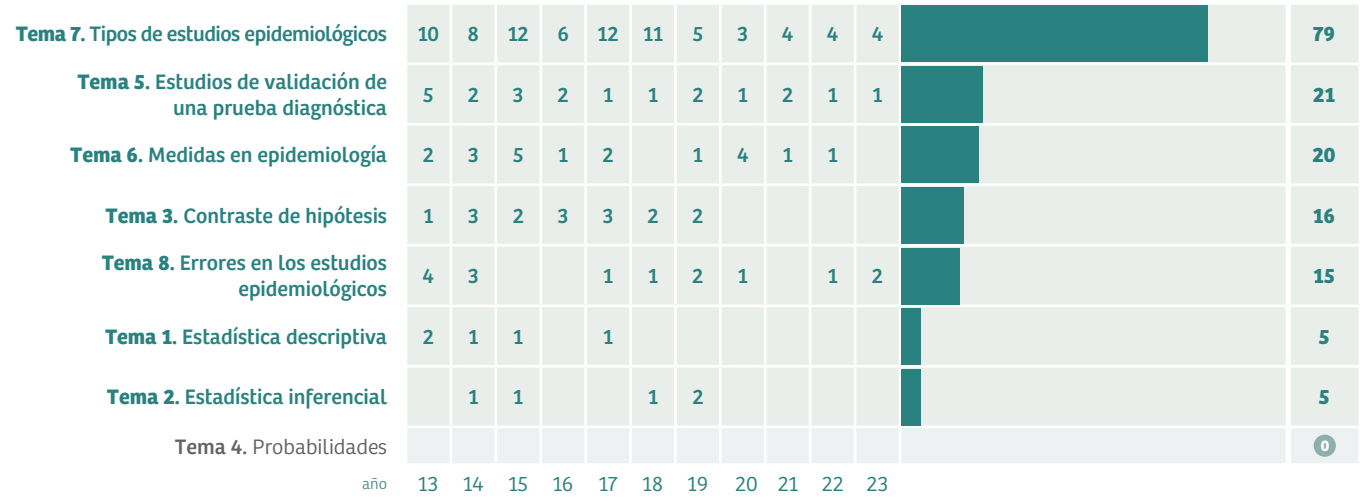
Tendencia general 2013-2023



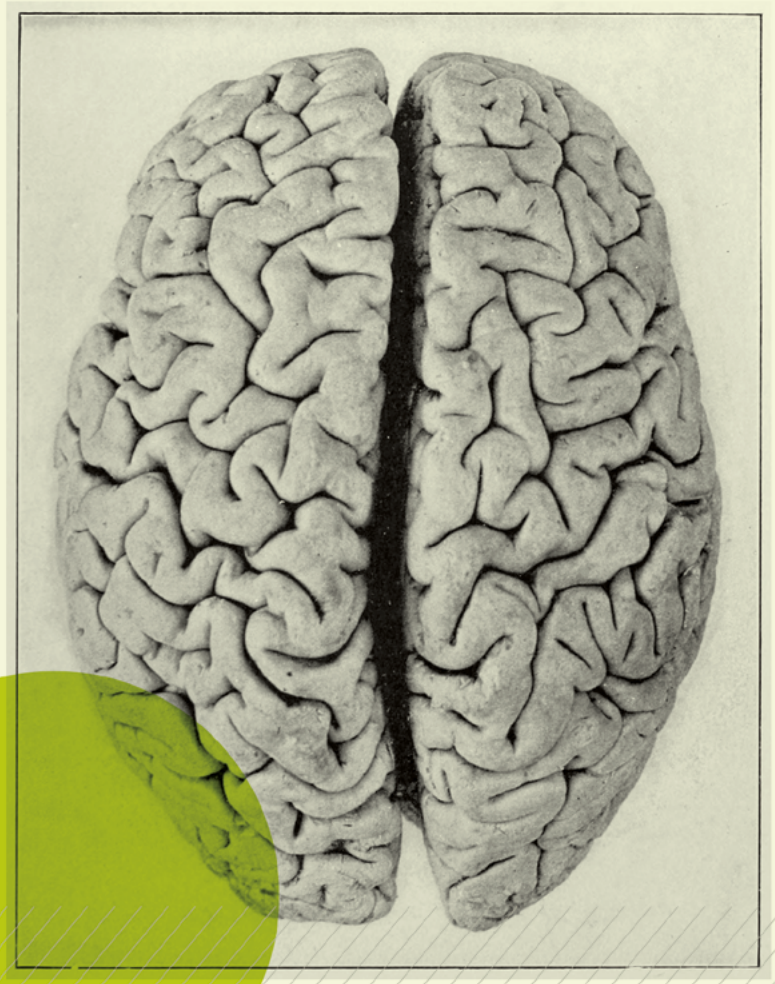
Importancia de la asignatura dentro del MIR



Distribución por temas



Índice



Curiosidad

Charles Spearman (Londres, 1863-1945), a quien hoy recordamos por el test de correlación de la "ρ" de Spearman, se dedicó fundamentalmente a lo largo de su vida al campo de la Psicología. Desarrolló la teoría bifactorial de la inteligencia (otra de sus aportaciones a la Estadística es el análisis factorial), por la cual existen dos factores que determinan la inteligencia de cada sujeto y que debían residir en partes distintas del cerebro: el factor G (genético y heredado), y el factor S (especial, que hace referencia a la capacidad concreta de cada sujeto para lidiar con cada problema específico).

Estadística

Tema 1

Estadística descriptiva

Autores: *Sergio Escribano Cruz (9), Carlos Corrales Benítez (16), Julio Sesma Romero (36), Víctor Rodríguez Domínguez (2).*

ENFOQUE MIR

Uno de los temas menos importantes de la asignatura, con 0-1 pregunta por término medio cada año. Lo más preguntado es el apartado de **técnicas de muestreo**. Lo siguiente en importancia son las propiedades de la **distribución normal**.En cuanto a las variables, es importante saber identificar cada tipo de **variable** pero son raras preguntas directas al respecto.

El objetivo de la Estadística es el estudio de una o varias características (variables) en una o varias poblaciones diana. Habitualmente el estudio de todos los individuos de dichas poblaciones es imposible por problemas logísticos, así que se suele estudiar sólo a un grupo reducido de individuos de cada población (muestra).

La **Estadística descriptiva** se ocupa de estudiar las variables que nos interesan de dicha muestra; como podemos estudiar a cada uno de los individuos de la muestra, todos los datos que obtengamos serán verídicos y no tendremos que extrapolar nuestros resultados, por lo que en Estadística descriptiva **no existe probabilidad de cometer errores**.

La **Estadística inferencial** intenta extrapolar cómo serían los resultados de la población objetivo si fuéramos capaces de estudiar a todos sus individuos. Para ello parte de los resultados obtenidos en la muestra. Así, los resultados estarán sujetos a una **probabilidad de error**, ya que si la muestra seleccionada no fuera representativa de la población, sus resultados no serían extrapolables a la misma.

Por último, el **contraste de hipótesis** compara los resultados de varias variables en una única población, o bien los resultados obtenidos para la misma variable en varias poblaciones. Al igual que en Estadística inferencial, para obtener los datos poblacionales se parte de resultados de las muestras estudiadas, por lo que existe **probabilidad de error**.

1.1. Técnicas de muestreo estadístico

El muestreo consiste en la selección de una muestra a partir de una población. El objetivo del muestreo es que la muestra escogida sea **representativa** de la población (esto es, que encierre toda la variabilidad posible que existe en la población), para que los resultados obtenidos en la muestra sean extrapolables a la población.

Antes de realizar la técnica de muestreo deseada, la **estratificación** nos puede ayudar a controlar una determinada variable que no queremos que influya en nuestros resultados **(MIR 12, 186)** para evitar que dicha variable actúe como factor de confusión **(ver tema 8. Errores en estudios epidemiológicos)**. La estratificación consiste en la división de la población en varias categorías según la variable mencionada, de modo que, una vez dividida la población, elegiremos sólo a individuos de entre las categorías de la variable que nos interese.

Ejemplo: nos interesa contrastar si el consumo de marihuana aumenta el riesgo de padecer esquizofrenia, pero no queremos que el consumo de otras drogas (posible factor de confusión) interfiera en nuestros resultados. Así, antes de escoger la muestra dividimos a la población en, por ejemplo, tres categorías en función de la variable "consumir otras drogas" (consumidores, no consumidores, exconsumidores), y posteriormente haremos el muestreo sólo en el grupo de no consumidores.

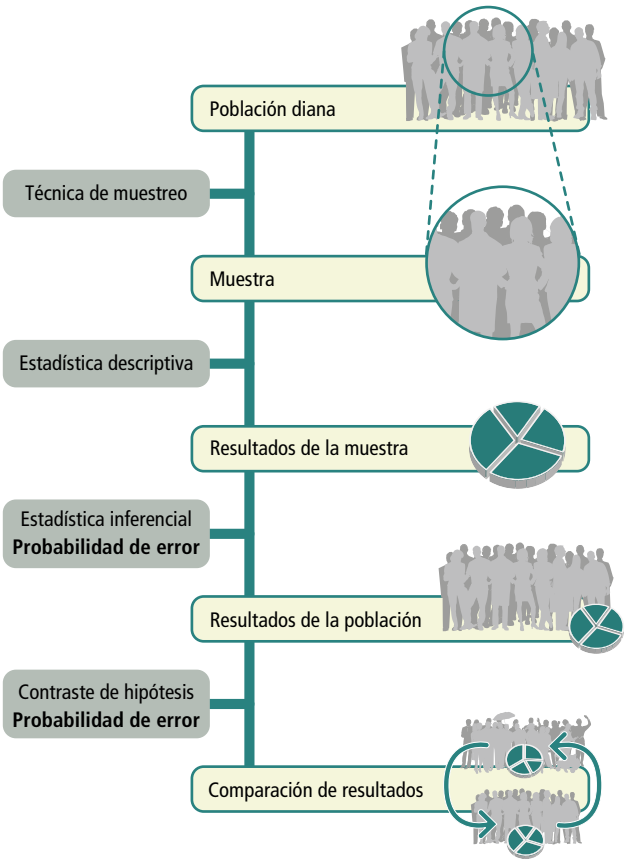


Figura 1. Esquema de realización de un estudio estadístico.

Técnicas de muestreo probabilístico

El muestreo probabilístico utiliza el **azar** para elegir la muestra de entre la población, lo cual permite conocer las probabilidades que tiene cada individuo de salir elegido. La utilización del azar para escoger la muestra (en lugar de cualquier criterio que defina el investigador) hace que existan más probabilidades de que la muestra sea representativa de la población, por lo que **las técnicas probabilísticas son mejores**.

Ejemplo: si de una población de 100 personas queremos coger 15 al azar, cada individuo tendrá 15/100 (15%) de probabilidades de salir escogido.

Muestreo aleatorio simple

Se asigna un número a cada individuo de la población, y posteriormente se escogen tantos números sean necesarios para completar el tamaño muestral requerido.

Ejemplo: para obtener una muestra de cinco individuos en una población de 100 personas, se asigna a cada persona un número del 1 al 100. Se introducen en una urna 100 pelotas numeradas, y se sacan de la urna cinco pelotas.

El muestreo aleatorio simple puede realizarse sin reposición de elementos (los individuos escogidos no pueden volver a ser elegidos) o con reposición de elementos (los individuos escogidos vuelven a ser introducidos en la población de la que se obtiene la muestra, de modo que podrían volver a salir elegidos). El muestreo con reposición de elementos es mejor porque se garantiza que en cada extracción de un individuo las probabilidades de salir elegido sean las mismas, pero en poblaciones pequeñas existirá el riesgo de que un mismo individuo salga elegido varias veces. Por tanto, el muestreo con reposición de elementos suele utilizarse en poblaciones grandes, donde la probabilidad de salir elegido dos veces es tan baja que el riesgo que se corre es pequeño.

Muestreo aleatorio sistemático

Se asigna un número a cada individuo de la población de manera aleatoria (en el muestreo aleatorio simple no hacía falta que esta asignación fuera aleatoria). Posteriormente, en vez de escoger "n" números, se escoge sólo uno, y a partir de él se obtiene el resto mediante una regla matemática.

Siempre y cuando se cumpla la premisa de ordenar a los individuos de la población inicialmente al azar, esta técnica es equivalente al muestreo aleatorio simple.

Ejemplo: para obtener una muestra de cinco individuos en una población de 100 personas, se asigna a cada persona, de forma aleatoria, un número del 1 al 100. Se introducen en una urna 100 pelotas numeradas, y la regla matemática va a ser "i + 10·x" (siendo "i" el número aleatorio obtenido, y "x" el número que va a ocupar cada individuo en nuestra muestra). Se saca una pelota de la urna y el número obtenido es el 17. Los individuos elegidos serán el 27, 37, 47, 57, 67.

Muestreo estratificado (MIR 17, 130)

Se denomina muestreo estratificado a aquel en el que, tras realizar estratificación de una determinada variable, se elige una muestra al azar de cada una de las categorías estudiadas de la variable.

Muestreo por conglomerados

Los conglomerados son grupos de individuos ya presentes de manera natural en la población y que encierran, en sí mismos, toda la variabilidad que posee la población diana. Son por tanto muestras perfectas que ya existen de manera natural. En el caso de identificar conglomerados en una población, se podría numerar a cada conglomerado y seleccionar, de manera aleatoria, el o los conglomerados necesarios.

En ocasiones estudiar un conglomerado entero puede resultar muy costoso por tener éste demasiado tamaño muestral. En ese caso podemos, dentro del conglomerado, realizar un muestreo aleatorio para seleccionar un menor número de individuos; como hemos realizado dos técnicas de muestreo una detrás de otra, este tipo de muestreo se llama **bietápico**.

Ejemplo: en una ciudad existen 10 hospitales que atienden un espectro de pacientes similar. Si queremos estudiar la población hospitalizada de dicha ciudad, en lugar de escoger una muestra de pacientes de los 10 hospitales, podríamos elegir al azar un único hospital (conglomerado) y estudiar a los pacientes ingresados en él.

Técnicas de muestreo no probabilístico

Los participantes en el estudio se seleccionan siguiendo criterios no aleatorios que define el investigador, por lo que, aunque se procura que la muestra sea representativa, las probabilidades de que no lo sea serán altas y la capacidad para extrapolar los resultados a la población será menor que con los métodos probabilísticos. Por lo tanto, son **peores que las técnicas probabilísticas**.

La técnica no probabilística más utilizada es el **muestreo de casos consecutivos**, que es la técnica de muestreo habitual de los ensayos clínicos.

Muestreo de casos consecutivos (MIR)

Consiste en reclutar a todos los individuos de la población accesible que cumplan los criterios de selección del estudio dentro de un intervalo de tiempo específico o hasta alcanzar un determinado número. Si se lleva a cabo de manera adecuada, la representatividad de la muestra puede ser semejante a la de un muestreo probabilístico.

Muestreo de conveniencia o accidental

Método sencillo y económico, que consiste en seleccionar sujetos accesibles, que estén a mano del investigador. Si el fenómeno estudiado no es suficientemente homogéneo en la población, las posibilidades de sesgo son muy elevadas.

Muestreo a criterio o intencional

En este tipo de muestreo el investigador incluye grupos de individuos que juzga típicos o representativos de la población, suponiendo que los errores en la selección se compensarán unos con otros.

1.2. Tipos de variables

Variables cualitativas (categóricas) (MIR 15, 184)

Hacen referencia a características que no se expresan mediante valores numéricos (p. ej., el color de pelo, la raza...).

Variables cualitativas ordinales (MIR)

Cuando los distintos valores de una variable cualitativa siguen un orden, nos interesará asignar a cada valor un número arbitrario (que nos inventamos) en función del orden que ocupa cada categoría. Esto es así porque los tests estadísticos que se utilizan para las variables que se expresan con números son más potentes que los tests empleados para variables cualitativas "puras".

Se distinguen de las variables cuantitativas en que los números asignados no cumplen propiedades matemáticas.

Ejemplo: escala del dolor: leve = 1, moderado = 2, intenso = 3. Tener un dolor "2" no significa tener el doble de dolor que un dolor "1".

Variables cualitativas nominales

Los valores de la variable no siguen un orden, y por tanto los nombraremos con palabras y no con números (p. ej., el color de pelo).

Cuando una variable cualitativa sólo puede tomar dos valores (p. ej., sexo: masculino o femenino) se denomina **dicotómica o binaria** (MIR). Si puede tomar más de dos valores se denomina **no dicotómica**.

Recuerda...

Las variables expresadas como **porcentajes** suelen ser variables **cualitativas**.
Ejemplo: si la prevalencia de EPOC es del 10%, la variable es tener o no tener EPOC, esto es, cualitativa.

Variables cuantitativas

Hacen referencia a características que se expresan mediante **valores numéricos** (p. ej., la tensión arterial, la temperatura...). Dichos valores numéricos cumplen las propiedades matemáticas de los números (p. ej., tener cuatro hijos implica tener el doble de hijos que una persona que tenga dos).

Variables cuantitativas discretas

Los valores numéricos no pueden adoptar cualquier valor (en general, sólo podrán ser números enteros).

Ejemplo: número de pacientes atendidos en un día en una consulta: se pueden atender 23 o 24 pacientes, pero no 23,5 pacientes. ¡Ojo! Al trabajar con estas variables, por ejemplo al calcular la media, sí podríamos obtener decimales.

Variables cuantitativas continuas

Los valores numéricos pueden adoptar cualquier valor, incluyendo decimales.

Ejemplo: presión arterial: si tuviera un aparato lo suficientemente preciso podría indicar una PAS de 140,6 mmHg. ¡Ojo! Aunque habitualmente sólo utilicemos una variable con números enteros, debemos pensar si sería posible dar un valor con decimales de dicha variable.

1.3. Medidas de análisis de los datos

Las variables **cualitativas** se suelen expresar mediante **porcentajes** (indicando el porcentaje de observaciones que presenta cada categoría de la variable), y no tienen medidas de dispersión.

Sin embargo, las variables **cuantitativas** se deben expresar mediante una medida de tendencia central y una medida de dispersión. Además, existen medidas de posición para indicarnos el lugar que ocupa cada observación dentro de la distribución.

Medidas de tendencia central

Informan acerca de cómo se agrupan los distintos valores registrados de los individuos de la muestra, indicando dónde se encuentra el centro de la distribución.

Media aritmética

La más utilizada, principalmente en distribuciones **simétricas**. Es el "centro de gravedad" del conjunto de valores. No debe usarse en distribuciones asimétricas ya que, al ser un cálculo matemático, los valores de los extremos influirán más que los centrales pudiendo artificialmente desplazar el valor de la media hacia ellos (en cuyo caso la media dejará de indicar dónde está el centro).

$$\bar{x} = \frac{\sum x_i}{n}$$

Mediana (MIR 11, 173; MIR)

Es el valor de la variable que presenta el individuo que ocupa la posición central si ordenamos las observaciones de menor a mayor, esto es, que divide el conjunto de observaciones en dos partes iguales (deja la mitad de las observaciones por encima y la mitad por debajo). Si la distribución de valores es simétrica, coincide con la media. Es la más indicada si los datos a analizar tienen una distribución **asimétrica** o presentan valores extremos.

Moda

Es el valor más repetido de todos los valores de la variable. Puede ser un valor único o haber varias. Es útil para distribuciones con varios "picos" de frecuencia, esto es, con varias modas.

Medidas de dispersión

Cuando analizamos los resultados, una variable cuantitativa en una muestra de sujetos, no sólo nos interesa en torno a qué valor se agrupan los resultados obtenidos (medida de tendencia central), sino también si las observaciones se encuentran "cerca" o "lejos" del centro de la distribución. Este dato lo indican las medidas de dispersión **(MIR 14, 190)**.

Para las variables de distribución **simétrica** se utiliza la **desviación típica**, y para las variables de distribución **asimétrica** el **rango intercuartílico**.

Ejemplo: la media de presión arterial sistólica de una muestra de pacientes puede ser de 130 mmHg porque la mitad tiene 129 mmHg y la otra mitad 131 mmHg (esta muestra tiene una PAS muy bien controlada), pero también puede ser 130 mmHg porque la mitad de pacientes tenga 90 mmHg y la otra mitad 170 mmHg (a pesar de tener la misma media, esta muestra es muy diferente de la otra, ya que los valores individuales están muy "alejados" del centro).

Las principales medidas de dispersión son **(MIR 10, 178)**:

Desviación típica (desviación estándar, σ)

Es la media de la diferencia que existe entre cada observación individual realizada y la media aritmética de la distribución. Se obtiene a partir de la raíz cuadrada de la **varianza (σ^2)**, que es la media del **cuadrado** de dichas diferencias.

$$\sigma^2 = \frac{\sum (x_i - x)^2}{n} \qquad \sigma = \sqrt{\sigma^2}$$

Para calcular la desviación típica es necesario realizar una argucia matemática, ya que si calculamos sin más la media de la diferencia o "separación" mencionada, al sumar la separación de los valores menores a la media (a la "izquierda"), que dará números negativos, más la separación de los valores mayores a la media (a la "derecha"), que dará números positivos, los números positivos se anularán con los negativos y obtendremos un resultado = 0.

*Dicha argucia matemática es la **varianza**, que es la media del **cuadrado** de la separación mencionada. Al elevar al cuadrado las separaciones "negativas", se vuelven números positivos y ya no se anulan con las separaciones positivas.*

Rango (recorrido)

Es la diferencia entre el valor máximo que toma la variable y su valor mínimo.

Rango intercuartílico

Es la diferencia entre el valor que ocupa el cuartil 3 (C3) de la distribución y el valor que ocupa el cuartil 1 (C1). Esto es, es el "rango" existente entre los individuos que se sitúan en el 50% central de la distribución.

Coeficiente de variación (MIR)

Se utiliza para comparar la dispersión de varias distribuciones, ya que **no tiene unidades** (es adimensional). Indica qué porcentaje respecto de la media supone la desviación típica de una distribución.

Ejemplo: no es lo mismo separarse (DT) 10 kg respecto a 50 kg de media (un 20% de separación) que respecto a 100 kg de media (un 10% de separación).

$$CV = \sigma / \bar{x}$$

¡Ojo! Cuando queremos expresar cualquier resultado en %, tenemos que multiplicar el resultado por 100, y viceversa, si queremos expresar un porcentaje en tanto por 1, deberemos dividir el resultado por 100.

Recuerda...

En variables cuantitativas de **distribución simétrica**, los resultados se expresan con la media y la desviación típica. En variables cuantitativas de **distribución asimétrica**, los resultados se expresan con la mediana y el rango intercuartílico.

Medidas de posición (localización)

Se basan en la ordenación de las observaciones de menor a mayor, y la posterior división de la distribución obtenida en grupos que contienen el mismo número de observaciones. A cada grupo se le asigna un número que indica el número de grupos situados a su "izquierda", esto es, que tienen valores de la variable **menores o iguales** a él. En general a estos grupos se les denomina "**centiles**", pero en función del número de grupos que se utilicen existen distintos nombres:

Cuartiles

Se divide a la distribución en cuatro partes iguales.

Deciles

Se divide a la distribución en 10 partes iguales.

Percentiles (MIR)

Se divide a la distribución en 100 partes iguales.

La **mediana** ocupa la posición central de una distribución, por lo que también es una medida de localización. Al situarse en el centro, equivale al cuartil 2 (C2), decil 5 (D5) o percentil 50 (p50).

*Ejemplo: el percentil 75 (p75) será el **valor** de la variable obtenido por aquél individuo tal que el 75% de las observaciones hayan sido menores o iguales a dicho valor, y el 25% de las observaciones hayan sido mayores a dicho valor. El p75 equivale al C3 y al D7,5.*

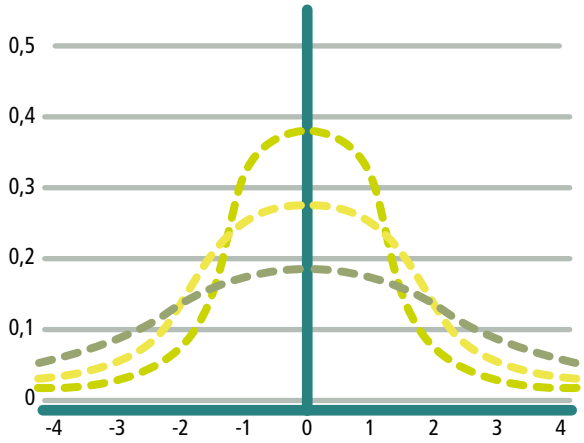


Figura 2. Dispersión de distribuciones.

Medidas de forma de una distribución

Cuando dos distribuciones coinciden en sus medidas de posición y dispersión, se hace difícil su comparación. Una manera de hacerlo es a través de la forma de la distribución. Para ello las distribuciones se comparan con la distribución normal en sus valores ideales, con media 0 y varianza 1 (distribución normal "tipificada"). Las dos medidas de la forma que se utilizan habitualmente son el grado de asimetría y el apuntamiento o curtosis.

Asimetría

Estudia la deformación horizontal de los valores en torno al valor central, la media, observando la concentración de la variable hacia uno de sus extremos. Se mide con los coeficientes de asimetría (el más utilizado es el coeficiente de asimetría de Fisher ó g1). Una distribución es simétrica cuando a la derecha y a la izquierda de la media existe el mismo número de valores, equidistantes dos a dos de la media, de tal manera que media, mediana y moda son iguales (g1 = 0).

Cuando tenemos una curva asimétrica a la izquierda o negativa, la mayoría de valores están a la derecha de la media (g1 <0), y la media es menor a la mediana, y ésta a su vez a la moda. Cuando tenemos una curva asimétrica a la derecha o positiva, la mayoría de valores se encuentra a la izquierda de la media (con g1 >0), y la media es mayor que la mediana, y ésta a su vez que la moda.

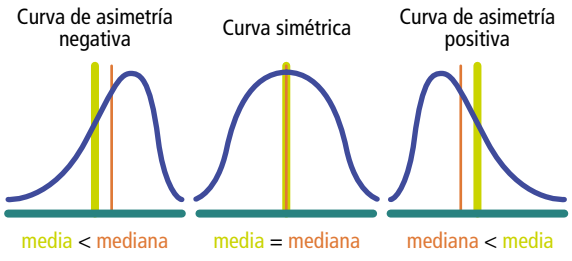


Figura 3. Asimetría.

Curtosis o apuntamiento (MIR 13, 177)

La curtosis mide el grado de agudeza o achatamiento de una distribución en relación a la distribución normal (determina cuán puntiaguda es una distribución). Se mide con el coeficiente de curtosis de Fisher (g2). Se dice que una curva es mesocúrtica cuando posee un grado de apuntamiento igual a la distribución normal (g2 = 0). Se denomina leptocúrtica si es más apuntada o puntiaguda (g2 >0). Se denomina platicúrtica si es más achatada (g2 <0).

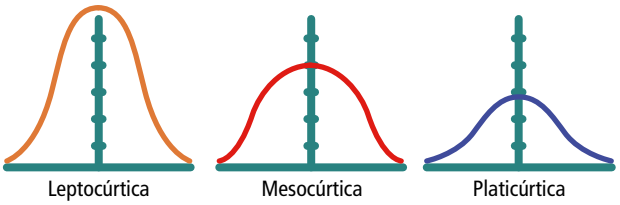


Figura 4. Curtosis.

Definición de una curva de distribución normal según la forma

Cuando una distribución de datos presenta un coeficiente de asimetría **g1 = ±0,5** y un coeficiente de curtosis de **g2 = ±0,5** cumple criterios de distribución normal.

1.4. Principales distribuciones de probabilidad

La "distribución" de los resultados de una variable es un modo de llamar a la morfología que toma la representación gráfica de dichos resultados. Cuando estudiamos los resultados de nuestro estudio, nos interesará que se distribuyan de forma similar a distribuciones ya conocidas y que tienen propiedades matemáticas interesantes, para que podamos aplicar dichas propiedades matemáticas a nuestros resultados.

Para las **variables cuantitativas continuas** nos interesará comprobar si se distribuyen de forma similar a la **distribución normal** (de Gauss).

Para las **variables cualitativas** y para las **cuantitativas discretas** podemos utilizar varias distribuciones, siendo las más utilizadas la **binomial** y la de **Poisson**.

Distribución normal (de Gauss) (MIR)

La mayoría de las **variables biológicas** (presión arterial, temperatura, datos de laboratorio, peso, altura, etc.) se distribuyen con este patrón.

Se define por una función de probabilidad continua, cuyo **rango** va desde –∞ hasta +∞, en la cual los valores se agrupan en torno a un valor central con forma de campana.

- Es simétrica.
- La media aritmética, mediana y moda coinciden (MIR).
- Es unimodal (tiene una única moda).
- El área bajo la curva de la distribución es igual a 1.

La distribución normal, aplicada a la estadística descriptiva representa el porcentaje de observaciones que tiene cada valor posible, por lo que la suma de todos los porcentajes (área bajo la curva) será = 100% = 1.

La principal utilidad matemática de la distribución normal es que permite definir una serie de **intervalos** que encierran un **área bajo la curva conocida**. En estadística descriptiva, esto implica que si nuestros resultados se distribuyen de un modo "normal", podremos establecer unos **intervalos** que indiquen **entre qué valores** se encuentra un determinado **porcentaje de las observaciones** de nuestra muestra **(MIR 13, 175; MIR)**:

- El intervalo $\bar{x} \pm \sigma$ comprende el 68% de los valores centrales u observaciones. Fuera de dicho intervalo queda el 32% de las observaciones (el 16% a cada lado).
- El intervalo $\bar{x} \pm 2\,\sigma$ comprende el 95% de los valores centrales u observaciones. Fuera de dicho intervalo queda el 5% de las observaciones (el 2,5% a cada lado).
- El intervalo $\bar{x} \pm 2{,}5\,\sigma$ comprende el 99% de los valores centrales u observaciones. Fuera de dicho intervalo queda el 1% de las observaciones (el 0,5% a cada lado).

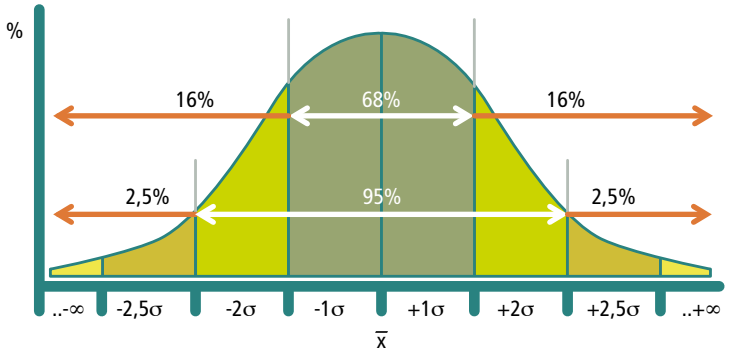


Figura 5. Distribución normal.

Distribución binomial

Se aplica a variables cuantitativas discretas o cualitativas, y consiste en convertir la variable en **dicotómica**, habiendo por tanto una probabilidad de "éxito" p(A) y una probabilidad de fracaso: su probabilidad complementaria p(1-A).

Distribución de Poisson

Es un caso particular de la distribución binomial que se utiliza para sucesos muy poco frecuentes: aquéllos en los que p(A) ó p(1-A) <10%, y además hay <5 individuos dentro de alguna categoría ($n \cdot p(A) < 5$ ó $n \cdot p(1\text{-}A) < 5$).

En la distribución de Poisson la media coincide con la varianza.

Tema 2

Estadística inferencial

Autores: *Sergio Escribano Cruz (9), Héctor Manjón Rubio (42), Julio Sesma Romero (36), Carlos Corrales Benítez (16).*

ENFOQUE MIR

La **inferencia de variables cuantitativas (medias)** era preguntada de forma repetida hasta hace años, a modo de problemas para calcular e interpretar intervalos de confianza ("existe un 95% de probabilidades de que el verdadero valor de la media se encuentre entre..."). Desde entonces ha habido menos preguntas al respecto, y han sido más teóricas. La **inferencia de variables cualitativas (porcentajes)** no la preguntan por lo que no la estudies.

Recuerda...

La **Estadística inferencial** estima cómo serían los resultados de la población objetivo si fuéramos capaces de estudiar a todos sus individuos. Para ello, extrae conclusiones a partir de los resultados obtenidos en la muestra, por lo que existirá una **probabilidad de error**.

2.1. Estadística inferencial para variables cuantitativas

El objetivo va a ser estimar, con un determinado nivel de **confianza**, entre qué niveles se encontrará la **verdadera media poblacional** de la variable que hemos medido en nuestra muestra.

Para ello pasamos de la distribución de resultados de nuestra muestra, que refleja el porcentaje o número de **observaciones** que tienen cada uno de los **valores** posibles de la variable, a una distribución de resultados poblacional, que refleja la **probabilidad** de que cada una de las posibles **medias** sea la verdadera media poblacional **(MIR 19, 123)**.

Teóricamente, la estadística inferencial simula qué ocurriría si, en vez de una sola muestra poblacional, fuéramos capaces de estudiar infinitas muestras poblacionales (que conjuntamente representarían a la población entera). De cada una de dichas muestras obtendríamos la media de la variable estudiada, y representaríamos dichas medias en una distribución de probabilidad **(ver figura 1)**. *El valor medio de las teóricas medias muestrales obtenidas es el valor más probable que adquirirá la verdadera media poblacional. En torno a dicho valor podemos construir intervalos (de confianza) que indicarán, con una cierta probabilidad (68%, 95% o 99%), entre qué valores se encontrará la verdadera media poblacional.*

Si la distribución muestral es normal, o si n >30 (teorema central del límite), la distribución poblacional también será normal y podremos utilizar las propiedades matemáticas de dicha distribución.

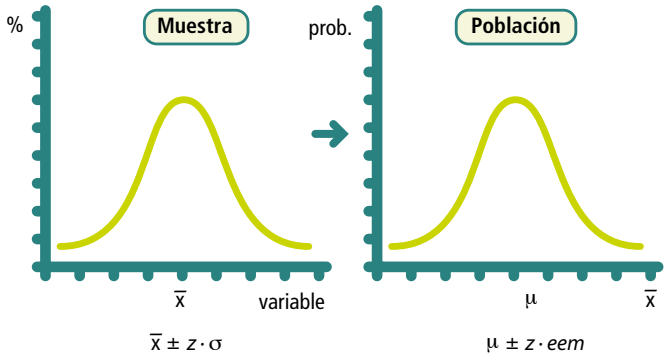


Figura 1. Estadística descriptiva (izquierda) y estadística inferencial (derecha).

Expresión de resultados de una inferencia de medias (MIR)

Al igual que al expresar los resultados de la muestra se utilizan **intervalos** que indican entre qué valores se encuentra un determinado porcentaje de las observaciones, al estimar los resultados de la población se utilizarán **intervalos de confianza (IC)** que indicarán entre qué valores se encuentra, con una determinada **probabilidad**, la verdadera media poblacional.

- La medida de tendencia central (**media poblacional** = μ) se equipara a la media muestral (μ = x).

  Si nuestra muestra es representativa de la población, la media muestral será el valor más probable que podrá tomar la media poblacional.

- La medida de dispersión utilizada se denomina **error estándar de la media** **(MIR 19, 122; MIR 12, 172)** (eem), y se calcula a partir de la desviación típica muestral.

$$eem = \frac{\sigma}{\sqrt{n}}$$

Para el cálculo de los **intervalos de confianza (IC)** se utilizan las propiedades matemáticas de la distribución normal **(MIR 15, 185)**:

- IC del 68% = μ ± eem
- IC del 95% = μ ± 2 eem
- IC del 99% = μ ± 2,5 eem

2.2. Estadística inferencial para variables cualitativas

El objetivo va a ser estimar, con un determinado nivel de **confianza**, entre qué niveles se encontrará el **verdadero porcentaje poblacional** de la categoría de la variable que hemos medido en nuestra muestra.

Para ello pasamos de una distribución de resultados binomial de nuestra muestra, que refleja el **porcentaje p(A)** de la categoría que queremos inferir y su porcentaje complementario p(1-A), a una distribución de resultados poblacional, que refleja la **probabilidad** de que cada uno de los posibles **porcentajes** sea el verdadero porcentaje poblacional. La variable de la distribución poblacional ("porcentaje poblacional") es **cuantitativa** y sigue una distribución normal.

¡Ojo! Al inferir un porcentaje, como empleamos la distribución binomial estamos realizando la estimación poblacional de una sola categoría de la variable (p. ej., en la variable "color de pelo", tendremos que elegir una sola categoría –pelo rubio, pelo castaño, pelo moreno...- cada vez que realicemos inferencia).

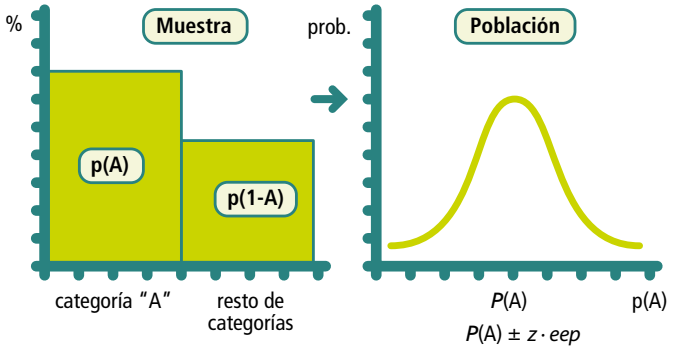


Figura 2. Estadística descriptiva (izquierda) y estadística inferencial (derecha).

Expresión de resultados de una inferencia de porcentajes

- La medida de tendencia central [**porcentaje poblacional** = P(A)] se equipara al porcentaje muestral [P(A) = p(A)].
- La medida de dispersión utilizada se denomina **error estándar del porcentaje** (eep). Como las variables cualitativas no tienen medidas de dispersión (no tienen desviación típica), se calcula a partir del porcentaje muestral.

$$eep = \sqrt{\frac{p(A) \cdot p\,(1\text{-}A)}{n}}$$

Para el cálculo de los **intervalos de confianza (IC)** se utilizan las propiedades matemáticas de la distribución normal:

- IC del 68% = P(A) ± eep
- IC del 95% = P(A) ± 2 eep
- IC del 99% = P(A) ± 2,5 eep

2.3. Cálculo del tamaño muestral para estudios de inferencia

Antes de realizar cualquier estudio epidemiológico, se debe analizar cuál es el tamaño muestral mínimo necesario para conseguir ofrecer unos resultados suficientemente precisos.

En los estudios de inferencia (**estimar cómo será un parámetro en la población**, p. ej., la prevalencia de una enfermedad) es necesario conocer los siguientes datos para calcular el tamaño muestral:

- Nivel de precisión (anchura del intervalo de confianza) deseado.
- Nivel de confianza deseado (95%, 99%...) **(MIR 18, 215)**; a menor nivel de confianza, menor amplitud del intervalo de confianza si mantenemos el mismo tamaño muestral.

Además, necesitamos otro dato que depende del **tipo de variable** utilizada en nuestro estudio:

- Variable cualitativa: **porcentaje** esperado del parámetro que se va a medir (según estudios previos) **(MIR)**.
- Variable cuantitativa: **varianza** de la variable **(MIR 14, 191)**.

No es necesario conocer **(MIR)**: **error beta**.

*Aclaración: en ocasiones, los autores de las preguntas MIR denominan erróneamente la probabilidad de error que existe en estadística inferencial (complementario del nivel de confianza del estudio) como **error alfa**. Sin embargo, debemos "aceptar" ese error como correcto cuando respondamos preguntas sobre el cálculo del tamaño muestral.*

Tema 3

Contraste de hipótesis

Autores: *Sergio Escribano Cruz (9), Víctor Rodríguez Domínguez (2), Carlos Corrales Benítez (16), Eduardo Franco Díez (5).*

ENFOQUE MIR

Es el tema más importante del bloque de Estadística del manual, si bien no han caído preguntas los últimos 3 años. Los conceptos más preguntados son los de error alfa y error beta, así como la interpretación de los resultados de un estudio en función del nivel de significación "p". Además, también pueden caer preguntas preguntas sobre los tests de contraste de hipótesis.

Recuerda...

El **contraste de hipótesis** compara los resultados de varias poblaciones entre sí, para lo cual debe realizar inferencia poblacional a partir de muestras obtenidas de cada población. Por tanto, al igual que en Estadística inferencial, existe **probabilidad de error**.

3.1. Errores en contraste de hipótesis

El contraste de hipótesis se utiliza en estudios que pretenden determinar si existen diferencias (comparación) o asociaciones (correlación) entre varias variables. El objetivo del contraste de hipótesis es determinar si esas diferencias o asociaciones observadas se deben al azar, o bien se deben a un efecto real (MIR).

Para ello, se definen dos hipótesis y las respectivas probabilidades de que cada una de ellas se deba al azar (errores alfa y beta).

- **Hipótesis nula (H_0)**: no existe asociación entre las variables analizadas.
- **Hipótesis alternativa (H_1)**: existe asociación entre las variables analizadas.

En la realidad sólo se podrá cumplir una de dichas hipótesis (o existe asociación, o no existe), pero al realizar nuestro estudio podemos acertar o bien equivocarnos, viendo asociación cuando no la hay (**error alfa**), o no viendo asociación cuando las hay en la realidad (**error beta**). Así, existen cuatro posibilidades si contrastamos los resultados de la realidad con los obtenidos en nuestro estudio (ver tabla 1).

Hipótesis nula y alternativa según el objetivo de nuestro estudio

Diseño de superioridad (MIR 19, 118; MIR)

El objetivo es conocer si una intervención "A" (tratamiento, prueba diagnóstica, etc.) es mejor que otra intervención "B", o bien si esa otra es mejor. En este caso H_0 es la igualdad "**A = B**", y H_1 es la presencia de **diferencias** "**A ≠ B**". Se utiliza por tanto un contraste de hipótesis **bilateral o de dos colas**, ya que nos interesa conocer si hay diferencias en ambos sentidos de la igualdad (A > B, B > A).

		REALIDAD	
		Se cumple H_1 ($A \neq B$)	Se cumple H_0 ($A = B$)
ESTUDIO TEST	Veo diferencias (se acepta H_1)	Potencia $1 - \beta$	Error Tipo I α
	No veo diferencias (no se rechaza H_0)	Error Tipo II β	$1 - \alpha$

Tabla 1. Contraste de hipótesis en estudios con diseño de superioridad.

Diseño de no inferioridad (MIR 18, 25; MIR 16, 29; MIR 16, 190; MIR 15, 190)

El objetivo es determinar si la intervención experimental "A" no es peor que otra ya existente "B"; nos da igual que sea igual o superior, lo que queremos es únicamente que no sea inferior. En este caso H_0 es la presencia de inferioridad "**A < B**", y la H_1 es la situación de **no inferioridad** "**A $\not<$ B**". Se utiliza por tanto un contraste de hipótesis **unilateral o de una cola**, ya que sólo nos interesa descartar que no haya diferencias en el sentido en que "A" es peor que "B" (A < B).

Para realizar un análisis de no inferioridad, debemos establecer un **límite de no inferioridad (δ = delta)** (MIR 10, 188) a partir del cual consideraremos que la intervención experimental es "inferior" a la ya existente. Dicho límite es arbitrario y suele establecerse en un 20% de diferencias: el fármaco experimental debe conseguir al menos el 80% del beneficio que consigue la intervención control.

Diseño de equivalencia terapéutica

El objetivo es determinar si la intervención experimental "A" es similar a otra ya existente "B"; la intervención experimental no debe ser mejor ni peor, sino producir un efecto terapéutico equivalente. En este caso H_0 es la ausencia de equivalencia "**A $\approx$ B**", y la H_1 es la situación de **equivalencia terapéutica** "**A ≈ B**".

Al igual que en un análisis de no inferioridad, debemos establecer unos límites arbitrarios para definir la situación de equivalencia. Dichos límites se suelen establecer en un

± 20%: el efecto de un fármaco debe encontrarse entre el 80% y el 120% del efecto que produce el otro (no puede ser más de un 20% peor ni más de un 20% mejor) **(MIR)**.

El ejemplo más típico de diseño de equivalencia terapéutica son los **estudios de bioequivalencia**, que se utilizan para autorizar la comercialización de los fármacos genéricos comparando sus propiedades farmacocinéticas con los respectivos fármacos originales **(ver tema 7. Tipos de estudios epidemiológicos)**.

Error tipo I (error alfa)
(MIR 13, 174; MIR 12, 173; MIR 11, 172; MIR 10, 176)

Es el error que se comete cuando las diferencias observadas se deben al azar (en la realidad, H_0 es cierta), pero el investigador lo interpreta como debido a una diferencia o asociación (en el estudio, se acepta H_1 y se rechaza H_0): es la **probabilidad de rechazar la hipótesis nula siendo cierta**. Por lo tanto, es un resultado "falso positivo".

La probabilidad de cometer este error es α, que define el **nivel de significación estadística** de los estudios epidemiológicos. Una vez realizado cualquier estudio epidemiológico de comparación, se calcula mediante un test estadístico el valor "**p**", que es la probabilidad de que una **diferencia igual o mayor** a la observada en el estudio no exista en la realidad (esto es, de que estemos incurriendo en un error α). Si el valor de "p" es inferior al nivel de significación estadística α que hayamos predefinido antes de iniciar el estudio (en general se define $\alpha = 0.05$), diremos que los resultados del estudio han sido **estadísticamente significativos** **(MIR)**.

- p <0,05: se acepta H_1 y se rechaza H_0.
- p >0,05: no se acepta H_1 y no se rechaza H_0.

El nivel de significación de un contraste de hipótesis es **independiente de la magnitud de las diferencias** encontradas entre las intervenciones que se comparan.

Regla mnemotécnica

Para recordar el error alfa: α-fetoproteína (α-**FP**).

El error tipo α es un resultado falso positivo (**FP**).

Error tipo II (error beta) (MIR)

Es el error que se comete cuando las diferencias observadas son reales (en la realidad, H_1 es cierta), pero el investigador lo interpreta como debido al azar (en el estudio, no se acepta H_1 y no se rechaza H_0): es la **probabilidad de no rechazar la hipótesis nula siendo falsa**. Por lo tanto, es un resultado "falso negativo".

Cuando se realizan estudios epidemiológicos y se concluye que no existen diferencias, se suele requerir una probabilidad de haber cometido un error beta <0,20 (menos del 20%) **(MIR 19, 115)**. *No obstante, el error beta es menos importante que el error alfa y en muchas ocasiones ni siquiera se calcula.*

Recuerda...

La hipótesis nula nunca se puede aceptar, y la hipótesis alternativa nunca se puede rechazar. Así pues:

- La hipótesis nula se rechaza o "no se rechaza".
- La hipótesis Alternativa se Acepta o "no se acepta".

Potencia estadística (poder estadístico)
(MIR 10, 179; MIR)

Es la probabilidad de detectar diferencias (en el estudio se acepta H_1 y se rechaza H_0) cuando en realidad existen (en la realidad, H_1 es cierta): es la **probabilidad de rechazar la hipótesis nula siendo falsa**. Por lo tanto, es un resultado "verdadero positivo".

La potencia estadística y el error beta son complementarios (potencia + $\beta = 1$). Por lo tanto:

Potencia estadística = $1 - \beta$
$\beta = 1 -$ potencia estadística

Así, de forma análoga al error beta, cuando se realizan estudios epidemiológicos y se concluye que no existen diferencias, se suele requerir que la potencia estadística sea al menos de un 80%.

Recuerda...

Los errores alfa y beta son **errores aleatorios**, esto es, debidos al azar (es el azar el que hace que el estudio falle y detecte diferencias cuando no las hay, o no las detecte cuando las hay). Los errores aleatorios se solucionan **aumentando el tamaño muestral**, por lo que ante un estudio cuyos resultados no sean estadísticamente significativos (p >0,05), si diseñamos un nuevo estudio incluyendo un mayor tamaño muestral, es posible que consigamos alcanzar entonces la significación estadística.

$\uparrow n \rightarrow \downarrow \alpha, \downarrow \beta, \uparrow$ potencia estadística

Recuerda...

Trucos para acertar las preguntas sobre contraste de hipótesis en el MIR:

- Las opciones categóricas ("siempre", "nunca", "sin lugar a dudas") son **falsas**. Se debe tener en cuenta que existe un margen de error que podemos cometer.
- Las opciones correctas suelen aplicar la definición de error alfa o error beta al ejemplo del enunciado, y para ello nos "traducen" la **tabla 1** de este tema. Son por ello opciones que parecen **trabalenguas** y que tienen el siguiente esquema con dos partes, la primera que nos habla de lo que ocurre en la realidad, y la segunda que nos habla sobre los resultados de nuestro estudio: *"En el caso de que no existieran diferencias entre los dos fármacos (= si en la realidad se cumple H_0), existiría una probabilidad de encontrar unos resultados como los obtenidos (= si, p. ej., en nuestro estudio hemos visto diferencias significativas -H_1-) inferiores al 5% (= hemos obtenido una p <0.05)".*

3.2. Cálculo del tamaño muestral en el contraste de hipótesis

Como en cualquier estudio epidemiológico, se debe analizar antes de comenzar el estudio cuál es el tamaño muestral mínimo necesario para conseguir unos resultados suficientemente precisos.

En los estudios de contraste de hipótesis (p. ej., comparar qué fármaco "A" o "B" es mejor) es necesario conocer los siguientes datos para calcular el tamaño muestral:

1. Aquellos parámetros que hacía falta conocer para estadística inferencial:
   - Nivel de precisión que queremos que tenga el intervalo de confianza.
   - Nivel de confianza deseado (68%, 95%, 99%).
   - Variabilidad del parámetro estudiado (según estudios previos), si la variable de interés es cuantitativa.
   - Porcentaje esperado del parámetro que se va a medir (según estudios previos), si la variable de interés es cualitativa.
2. Parámetros específicos del contraste de hipótesis:
   - Tipo de diseño del estudio y si el análisis será de una cola o de dos colas.
   - Error tipo α y tipo β permitidos: nivel de potencia estadística deseado. Cuanta mayor potencia, y cuanto menor α y β deseados, mayor tamaño muestral **(MIR 10, 190)**.
   - Magnitud de la diferencia mínima clínicamente relevante que se desea demostrar entre los dos fármacos (δ).

     Aclaración: se llama también delta, pero es un concepto distinto al límite de no inferioridad.
   - Porcentaje de pérdidas previsto (**d**) **(MIR)**.

No es necesario conocer: nivel de **enmascaramiento** del estudio **(MIR)**, número de pacientes que somos capaces de reunir **(MIR)**, número de centros participantes **(MIR)**.

Si al finalizar el estudio se obtiene un resultado no significativo, no se deben añadir pacientes hasta que lo sea, sino revisar la hipótesis de trabajo y la determinación del tamaño muestral y realizar un nuevo estudio **(MIR)**.

3.3. Tests para contraste de hipótesis

Tests para estudios de comparación de variables

Los principales tests para comparación de variables se exponen en la **tabla 2**. Para elegir el tipo de test a utilizar nos deberemos fijar en dos criterios fundamentales:

- Qué tipo de variable (**cualitativa o cuantitativa**) es la **variable resultado** que tenemos que comparar. Los tests para variables cuantitativas aportan una mayor potencia estadística (permiten alcanzar la significación estadística con menor tamaño muestral y sus resultados son más precisos) que los utilizados para variables cualitativas.

  Cuando la variable es **cuantitativa**, además, tendremos que elegir entre los siguientes grupos de tests estadísticos:
  - **Tests paramétricos**: se utilizan cuando la variable sigue una distribución normal **(MIR)**, o bien si n >30 (pese a que la distribución no sea normal). Aportan una mayor potencia estadística que los no paramétricos.
  - **Tests no paramétricos**: se utilizan cuando la variable no sigue una distribución normal y además n <30.

  Las variables **ordinales** se consideran como si fueran **cuantitativas**, pero con la restricción de que sólo se puede emplear con ellas **tests no paramétricos (MIR)**.
- Si estamos comparando entre sí los resultados obtenidos en esa variable en **varios grupos de individuos (datos independientes)**, o bien en un único grupo de individuos pero en **varios momentos del tiempo (datos apareados)**.

	VARIABLE CUALITATIVA		VARIABLE CUANTITATIVA	
			2 GRUPOS O 2 MOMENTOS DEL t	>2 GRUPOS O >2 MOMENTOS DEL t
DATOS INDEPENDIENTES (VARIOS GRUPOS)	chi^2 (χ^2) • Corrección de Yates* • Test exacto de Fisher**	**PARAMÉTRICO**	t Student	ANOVA
		NO PARAMÉTRICO V. ORDINALES	Mann-Whitney	Kruskal-Wallis
DATOS APAREADOS (VARIOS MOMENTOS DEL t)	McNemar	**PARAMÉTRICO**	t Student para datos apareados	ANOVA para datos apareados
		NO PARAMÉTRICO V. ORDINALES	Wilcoxon	Friedmann

*Corrección de Yates: corrección que se aplica al test de chi^2 cuando el tamaño muestral es n <200.

**Test exacto de Fisher: cuando en la tabla de contingencia de la chi^2 hay menos de cinco individuos en >25% de las casillas [expresado matemáticamente: $n \cdot p < 5$ ó $n \cdot (1-p) < 5$] no se puede utilizar el test de chi^2 y hay que utilizar el test exacto de Fisher.

Tabla 2. Tests de contraste de hipótesis para comparación de variables **(MIR 17, 122; MIR 17, 123; MIR 17, 124; MIR 10, 177)**.

Recuerda...

Las variables resultado cualitativas nos las plantearán habitualmente como porcentajes (comparar varios porcentajes), mientras que las variables resultado cuantitativas nos las plantearán habitualmente como medias (comparar varias medias).

Regla mnemotécnica

Test de contraste de hipótesis para variables cualitativas

CHI** tuviera un **YATE** iría a **PESCAR** a **NEMO

Datos independientes:
CHI cuadrado
Corrección de **YATE**s
Corrección de **Fisher (PESCAR)**

Datos apareados:
Test de Mc**NEMAR**

Tests para estudios de asociación entre variables

En este caso, lo que se pretende es demostrar si los cambios que se produzcan en una o varias variables (variables independientes, x_i) van a influir sobre el valor que tome otra variable (variable dependiente, y); además, se pretende **cuantificar** dicha influencia. Todas las variables se recogen de **una misma muestra**.

Regresión

La **regresión** trata de expresar mediante ecuaciones la asociación existente (mostrar mediante una fórmula matemática cómo varía la variable "y" con cada unidad de aumento de las variables "x_i"). Además, las ecuaciones obtenidas nos permitirán **predecir** el valor que tomará la variable "y" en un individuo para el que conocemos las variables "x_i". Las variables introducidas pueden ser tanto cuantitativas como cualitativas (en cuyo caso habrá que asignar a cada categoría un número que nos inventemos).

Por ejemplo: en una muestra de individuos, analizar cuánto aumenta el colesterol (variable y) con cada kg que aumente el peso medio (variable x) en dicha muestra.

Si existe sólo una variable independiente (x_i) en la ecuación se habla de **regresión univariante o simple**. Si existen dos o más variables independientes (x_i) en la ecuación se habla de **regresión multivariante o múltiple** (MIR 16, 194). Si se utiliza regresión multivariante, todas las variables independientes incluidas en la ecuación quedan "ajustadas entre sí" de modo que el coeficiente que acompaña a cada variable indicará el efecto que tiene exclusivamente dicha variable sobre la variable "y", eliminando el efecto de cualquier otra variable independiente introducida en la ecuación: sirve por tanto para **evitar sesgos por factor de confusión**.

El tipo de variable de la variable dependiente (y) define el tipo de regresión:

- **Regresión logística**: si la variable "y" es cualitativa (MIR 12, 176; MIR).
- **Regresión lineal**: cuando la variable "y" es cuantitativa, la fórmula matemática más empleada es la ecuación de una recta:

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + ... + b_i \cdot x_i$$

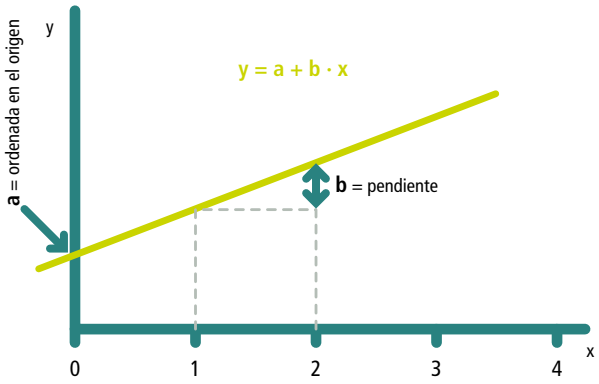


Figura 1. Regresión lineal simple.

El valor de la constante "**a**" indica el valor que toma la variable "y" (eje de ordenadas) cuando las variables independientes valen = 0. Se denomina **ordenada en el origen**.

El valor de cada coeficiente "**b**" expresa cuantitativamente la asociación entre cada variable "x_i" y la variable "y": indica cuánto aumenta la variable "y" con cada unidad de aumento de cada variable "x_i" (MIR). Se denomina **pendiente**.

Regresión de Cox: método de regresión que se utiliza en el análisis de supervivencia.

Correlación (MIR 15, 186)

La **correlación** trata de expresar, mediante un **coeficiente de correlación**, el porcentaje de los cambios observados en la variable dependiente que se deben a los cambios observados en las variables independientes. Por lo tanto, indicará lo "fuerte" que es el grado de asociación.

Evidentemente, los cambios que ocurran en una muestra de pacientes en la variable "y" (p. ej., en el colesterol), no se deberán en su totalidad a los cambios apreciados en la variable "x" (p. ej., el peso). Sólo un cierto porcentaje de esa variación se deberá a la variable "x", y el resto se deberá a otras variables que no estamos estudiando (p. ej., la dieta, la realización o no de ejercicio físico, etc.).

Los **tests de correlación** más utilizados son los empleados para evaluar la correlación existente entre **dos variables cuantitativas**.

- **Coeficiente "r" de Pearson** (MIR): es un test paramétrico que mide el grado de correlación lineal entre las variables (se emplea cuando las dos variables siguen una distribución normal o bien si n >30). No descarta otros tipos de correlación que no sea la lineal.
- **Coeficiente "ρ" de Spearman**: es un test no paramétrico (se emplea cuando alguna de las variables sigue una distribución no normal y además n <30).

El **signo** del coeficiente de correlación (+/−) indica si la correlación es positiva (cuando la variable "x" aumenta, la variable "y" aumenta) o si es negativa (cuando la variable "x" aumenta, la variable "y" disminuye).

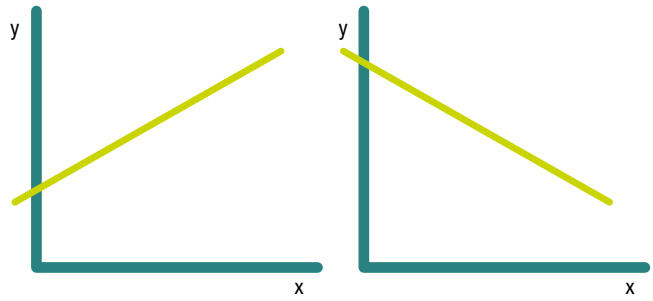


Figura 2. Correlación positiva (izquierda) y negativa (derecha).

El **valor absoluto** del coeficiente indica, si lo elevamos al cuadrado, el porcentaje de los cambios de la variable "y" que se explican por los cambios de la variable "x" (p. ej., un coeficiente de 0,8 = 80%, indica que el 64% de los cambios en la variable "y" se explican por los cambios en la variable "x"):

- Valor absoluto >0,7: correlación fuerte **(MIR 14, 192)**.
- Valor absoluto <0,7: correlación débil.
- Valor absoluto = 0: ausencia de correlación.

Análisis de supervivencia (MIR 14, 33; MIR)

Se utiliza cuando en un estudio epidemiológico la variable respuesta es el **tiempo que transcurre hasta que sucede un evento de interés** (la muerte, la aparición de enfermedad, la curación, el alta hospitalaria...). Así pues, las variables tienen una parte cuantitativa (tiempo que transcurre) y una parte cualitativa (aparición o no de un evento).

Cuando el tiempo de seguimiento de alguno de los pacientes del estudio termina antes de que haya tenido lugar el evento de interés se habla de **observaciones incompletas o censuradas**. Si un paciente fallece por una causa distinta a la enfermedad estudiada se considerará como censurado, ya que, en caso contrario, se estaría cometiendo un sesgo de información.

En la representación gráfica de las curvas de supervivencia, se suele anotar al principio de cada unidad de tiempo los pacientes que siguen en el estudio y todavía no han presentado el evento de interés (**pacientes en riesgo**). Para calcular los pacientes en riesgo al inicio de cada unidad de tiempo, se deben eliminar tanto los pacientes que han tenido el evento de interés como los pacientes censurados.

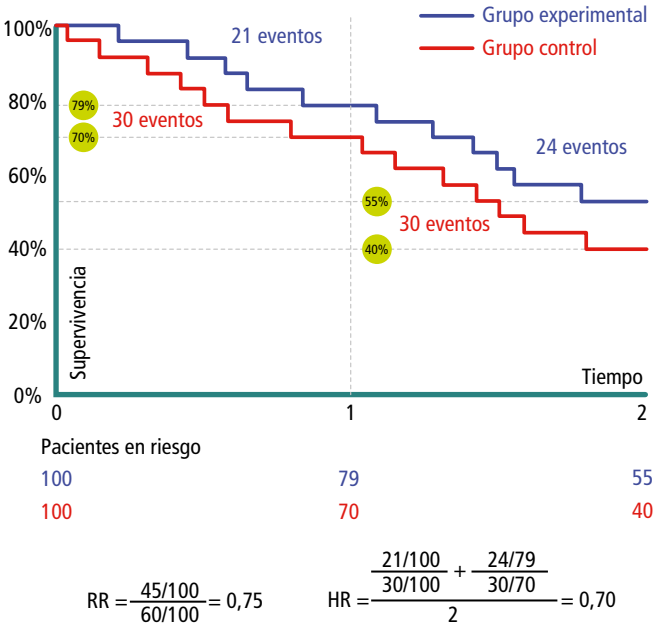
Los métodos estadísticos **no paramétricos** son los más frecuentemente utilizados en análisis de supervivencia. Entre ellos los más destacados son:

- Kaplan-Meier: utilizado para "calcular" las curvas de supervivencia **(MIR 18, 214; MIR 12, 174)**.
- Test de log-rank: utilizado como test de comparación, es similar al chi^2 (comparar los resultados obtenidos entre varias intervenciones).
- Modelo de regresión de Cox: utilizado para realizar regresión.

Para cuantificar el **grado de asociación** existente entre un determinado factor de riesgo o protector y un evento de interés estudiado con análisis de supervivencia, la medida epidemiológica utilizada es el **hazard ratio** o razón de riesgos (HR). Su interpretación es similar a las del resto de medidas de asociación (RR, OR...).

El HR es el cociente entre el riesgo que tiene de presentar el evento de interés un sujeto del grupo experimental respecto a un sujeto del grupo control, **por cada unidad de tiempo** que dura el estudio **(MIR 14, 34)**. Es similar al riesgo relativo (RR), dado que también es un cociente de riesgos. Sin embargo, mientras el RR compara el riesgo acumulado **a lo largo de todo el estudio** (cociente de incidencias acumuladas al finalizar el estudio), el HR analiza el **riesgo instantáneo** para cada unidad de tiempo (cociente entre la **velocidad** de progresión de la enfermedad o "hazard rate" de los grupos comparados). Así, el HR analiza las probabilidades de presentar el evento en el siguiente instante de tiempo, para aquellos individuos que continúen en el estudio al inicio de dicho periodo de tiempo (pacientes en riesgo); el RR analiza las probabilidades de presentar el evento a lo largo de todo el estudio.

Ejemplo **(ver figura 3)***: imaginemos un estudio que compara 2 grupos de 100 pacientes, que dura 2 unidades de tiempo, y que tiene un HR de 0.7 (sin pérdidas). Pongamos que observamos, por ejemplo, 30 eventos en el grupo control en cada periodo de tiempo. En este caso, en el grupo experimental habría 21 eventos en el periodo de tiempo 1 (un 70% de 30) y quedarían 79 pacientes para el periodo de tiempo 2. En dicho periodo de tiempo habría 24 eventos (en el grupo control hay 30 eventos de 70 pacientes que quedan, esto es, un riesgo del 42,8%; el riesgo del grupo experimental debe ser el 70% de ese 42,8%: un 30% sobre 79 pacientes, que son 24 eventos). El HR del estudio es 0.7, pero el RR sería igual al cociente de incidencias acumuladas: 45 eventos en el grupo experimental / 60 eventos en el grupo control = 0.75. Así, vemos que el HR y el RR son similares, pero no son la misma cosa.*



$$RR = \frac{45/100}{60/100} = 0{,}75 \qquad HR = \frac{\frac{21/100}{30/100} + \frac{24/79}{30/70}}{2} = 0{,}70$$

Figura 3. Curvas de Kaplan-Meier que representan el ejemplo expuesto en el texto.

Tema 4

Probabilidades

Autores: *Sergio Escribano Cruz (9), Julio Sesma Romero (36), Víctor Rodríguez Domínguez (2), Héctor Manjón Rubio (42).*

ENFOQUE MIR

Tema no preguntado en el MIR desde hace más de 15 años. Por lo tanto, **no lo estudies**.

Conceptos

- **Probabilidad**: medida de la verosimilitud de que un determinado suceso ocurra o no. Oscila entre 0 (suceso imposible) y 1 (suceso seguro).
- **Sucesos complementarios**: dos sucesos A y B son complementarios cuando la suma de las probabilidades de ambos es igual a 1. Siempre que no ocurre un suceso, ocurre el suceso contrario: p(A) + p(B) = 1.

  Ejemplo: ser hombre (A) y ser mujer (B).
- **Sucesos incompatibles**: se denomina así a los sucesos excluyentes, es decir, que no pueden suceder a la vez. Dos sucesos A y B son incompatibles cuando p(A ∩ B) = 0

  Ejemplo: tener el pelo moreno (A) o pelirrojo (B).
- **Sucesos independientes**: la probabilidad de que ocurra uno de ellos no se influye por el hecho de que ocurra o no el otro: p(A/B) = p(A); p(B/A) = p(B).

  Ejemplo: ganar la quiniela (A) y ganar la lotería (B).

Unión de probabilidades (∪)

Es la probabilidad de que ocurra un suceso u otro. Al calcular la unión de probabilidades se suma la probabilidad de que ocurra cada suceso, pero se debe restar una vez la probabilidad de que ocurran ambos a la vez (ya que al sumar la probabilidad de que ocurra cada suceso se está contando dos veces a los individuos que presentan los dos sucesos):

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

Si tenemos dos **sucesos incompatibles**: p(A ∩ B) = 0, y por tanto p(A ∪ B) = p(A) + p(B)

Si queremos calcular la probabilidad de que sólo ocurra un suceso u otro (eliminando por tanto todos los casos en los que aparezcan los dos sucesos a la vez) debemos restar dos veces en la fórmula la intersección de probabilidades:

$$p(\text{sólo A ó B}) = p(A) + p(B) - 2 \cdot p(A \cap B)$$

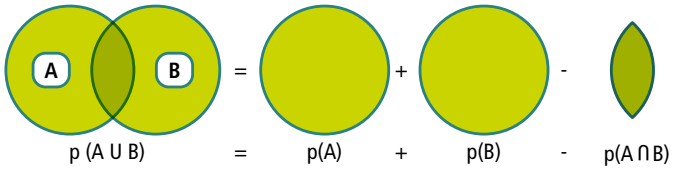


Figura 1. Unión de probabilidades.

Intersección de probabilidades (∩) (MIR)

Es la probabilidad de que ocurran un suceso **y** otro simultáneamente. Para calcularlo se multiplica la probabilidad de que ocurra uno de ellos [p(A)] por la probabilidad de que ocurra el otro en aquellos casos en los que ocurre el primer suceso [probabilidad condicionada = p(B/A)]:

$$p(A \cap B) = p(A) \cdot p(B/A) = p(B) \cdot p(A/B)$$

Si tenemos dos **sucesos independientes**: p(B/A) = p(B), y por tanto p(A ∩ B) = p(A) · p(B).

Probabilidad condicionada

Una probabilidad de un suceso A condicionada al suceso B es la probabilidad de que ocurra el suceso A considerando sólo los casos en los que ocurre B (es decir, la probabilidad de que ocurra A sabiendo que ha ocurrido B).

La fórmula por la cual se puede calcular la probabilidad condicionada p(A/B) a partir de la probabilidad condicionada p(B/A) se denomina **teorema de Bayes**.

$$p(A/B) = \frac{p(A \cap B)}{p(B)} = \frac{p(A) \cdot p(B/A)}{p(B)}$$

A menos que tengamos sucesos independientes, en cuyo caso las fórmulas se simplifican, no nos van a poder pedir en el MIR calcular la probabilidad condicionada ni la intersección de probabilidades.

Epidemiología

Tema 5

Estudios de validación de una prueba diagnóstica

Autores: *Víctor Rodríguez Domínguez (2), Rafael Leal Hidalgo (47), Héctor Manjón Rubio (42), Eduardo Franco Díez (5).*

ENFOQUE MIR

Tras el tema de Tipos de estudios epidemiológicos (el más importante), el de Estudios de evaluación de una prueba diagnóstica y Medidas en Epidemiología son el segundo y tercero en importancia, respectivamente.
De validación de pruebas diagnósticas suelen hacer **1-3 preguntas en los últimos años**. Hacen siempre alguna pregunta teórica sobre los conceptos de sensibilidad, especificidad, valor predictivo positivo y valor predictivo negativo. Además, pueden caer problemas para calcular esos conceptos. En los últimos años han preguntado por las **razones de verosimilitud**.

Cuando se desea comercializar un nuevo test diagnóstico (p. ej., un nuevo modelo de esfingomanómetro), se deben llevar a cabo **estudios de validación** mediante los cuales se evaluarán distintas cualidades del test:

Validez (exactitud)

Es el grado en el cual una medición representa el verdadero valor que se desea medir. En los estudios de validación, representaría el grado de correlación de las medidas obtenidas mediante el test con las obtenidas mediante el *gold standard* **(MIR)**.

Reproducibilidad (fiabilidad, precisión)

Es la capacidad del test de obtener el mismo resultado cuando la medición se repite bajo las mismas condiciones de medida.

Concordancia

Es la capacidad del test de obtener el mismo resultado cuando la medición se repite mediante **distintas condiciones** de medida (p. ej., cuando la persona encargada de realizar el test es distinta). El cambio en condiciones que afectan a la validez externa de una prueba (como la prevalencia de enfermedad, o la aplicación del test como *screening* o como diagnóstico de confirmación) afecta al grado de concordancia existente.

Los estudios de concordancia utilizan distintos tests estadísticos en función de cómo sea la variable resultado que se va a utilizar:

- Variable cualitativa dicotómica: **estadístico kappa** (de Cohen) **(MIR 13, 176)**. Oscila entre −1 (excesiva discordancia) y +1 (concordancia completa). Cuando es igual a 0, la concordancia obtenida se debe al azar.

  Ejemplo: evaluar la concordancia entre dos radiólogos a los que se les muestran las mismas radiografías de tórax y tienen que indicar si hay SÍ/NO un infiltrado neumónico.

- Variable cualitativa no dicotómica: **estadístico kappa ponderado**. Es igual que el estadístico kappa, pero tiene en cuenta el grado de discordancia existente, lo cual es importante cuando existen varias categorías posibles de la variable (por eso se usa en variables no dicotómicas).

  Cuantas más categorías posibles tenga una variable cualitativa, más difícil va a ser que dos observadores distintos indiquen exactamente la misma categoría ante una misma muestra. Por lo tanto, si usamos el estadístico kappa, cuantas más categorías existan, menos grado de concordancia calcularemos. Es por eso que en variables con varias categorías (no dicotómicas) se emplea el test de kappa ponderado.

  Ejemplo: evaluar la concordancia entre dos cardiólogos que definen la clase funcional de la NYHA I-II-III-IV de una serie de pacientes. Existirá más concordancia si cuando un cardiólogo indica clase II el otro indica clase III, que si un cardiólogo indica clase I y el otro clase IV.

- Variable cuantitativa: **coeficiente de correlación intraclase** **(MIR 11, 175)**.

  Ejemplo: evaluar la concordancia entre dos anatomopatólogos que cuantifican el número de mitosis en una misma serie de muestras de biopsias de un tumor neuroendocrino.

5.1. Parámetros de validez de una prueba diagnóstica

Para evaluar la validez de una prueba diagnóstica, se realiza un **estudio transversal** mediante el cual se comparan los resultados obtenidos por el test (que cataloga a los individuos en "+" o "−") con los resultados obtenidos por el mejor método diagnóstico que esté disponible, llamado ***gold standard*** o **patrón oro** (que va a catalogar a los individuos del estudio en "enfermos" o "sanos") **(MIR)**.

Dicho estudio debe realizarse en las condiciones más similares posibles a la práctica clínica habitual. Además, la comparación debe ser ciega e independiente y abarcar todo el espectro de la enfermedad (MIR).

	GOLD STANDARD		
	ENFERMOS	**SANOS**	
TEST POSITIVO	VP	FP	VP + FP total de positivos
TEST NEGATIVO	FN	VN	FN + VN total de negativos
	VP + FN total de enfermos	FP + VN total de sanos	n

Tabla 1. Estudio de validación de una prueba diagnóstica (MIR 10, 196).

Parámetros de validez interna

La **validez interna** es la capacidad del test de obtener resultados exactos (que representen el verdadero valor que se desea medir) en los sujetos de la **muestra** que se ha utilizado para realizar el estudio.

Los parámetros de validez interna son características intrínsecas del test que no dependen de la población a la que se aplique (esto es, **no dependen de la prevalencia de enfermedad**) (MIR 16, 206).

Sensibilidad (S) (MIR 15, 235; MIR 12, 194; MIR 11, 189)

Es la capacidad del test de detectar a los sujetos enfermos. Es la probabilidad de que un sujeto enfermo (según el *gold standard*) saque "+" en el test. La probabilidad complementaria a la sensibilidad (esto es, la probabilidad de que un sujeto enfermo saque "–" en vez de "+" en el test) es la **tasa de falsos negativos (TFN)** (MIR 13, 198).

S = VP / total de enfermos
TFN = FN / total de enfermos
S + TFN = 1 → S = 1 – TFN; TFN = 1 – S

Así, un test muy sensible es útil en la práctica cuando su **resultado es negativo** (MIR), ya que el test tendrá una TFN muy baja y por lo tanto casi todos los pacientes negativos serán verdaderos negativos (sanos), pudiendo por tanto **descartar enfermedad**.

La sensibilidad es análoga a la **potencia estadística** de un estudio de contraste de hipótesis (MIR).

Especificidad (E) (MIR 19, 131; MIR)

Es la capacidad del test de detectar a los sujetos sanos. Es la probabilidad de que un sujeto sano (según el *gold standard*) saque "–" en el test (MIR 22, 44). La probabilidad complementaria a la especificidad (esto es, la probabilidad de que un sujeto sano saque "+" en vez de "–" en el test) es la **tasa de falsos positivos (TFP)**.

E = VN / total de sanos
TFP = FP / total de sanos
E + TFP = 1 → E = 1 – TFP; TFP = 1 – E

Un test muy específico es útil en la práctica cuando su **resultado es positivo**, ya que el test tendrá una TFP muy baja y por lo tanto casi todos los pacientes positivos serán verdaderos positivos (enfermos), pudiendo por tanto **confirmar enfermedad**.

Razón o cociente de probabilidad (razón de verosimilitud, *likelyhood ratio*, índice de eficiencia pronóstica)

Las razones de verosimilitud dividen las probabilidades de que un individuo enfermo y un individuo sano tengan resultados positivos (razón de probabilidad positiva) o negativos (razón de probabilidad negativa) en una prueba diagnóstica. Así, indican cuántas veces es más probable que un enfermo obtenga un resultado determinado (positivo o negativo) respecto a un individuo sano. Dan información, por lo tanto, de cuánto se modifica la probabilidad pre-test de enfermedad al obtener el resultado (positivo o negativo) de la prueba diagnóstica.

La razón de probabilidad positiva (RPP, RVP) (MIR 21, 52; MIR 16, 205; MIR 13, 196) es el cociente entre la probabilidad de que un enfermo obtenga un resultado positivo (S) y la probabilidad de que un sano obtenga un resultado positivo (TFP). Cuanto **mayor** sea dicha razón (más probabilidades de que un enfermo sea positivo respecto de que lo sea un sano), mejor será la prueba.

La razón de probabilidad negativa (RPN, RVN) es el cociente entre la probabilidad de que un enfermo obtenga un resultado negativo (TFN) y la probabilidad de que un sano obtenga un resultado negativo (E). Cuanto **menor** sea dicha razón (menos probabilidades de que un enfermo sea negativo respecto de que lo sea un sano), mejor será la prueba.

RPP = S / TFP
RPN = TFN / E

La capacidad diagnóstica de un test se puede clasificar en función del valor numérico de la razón de probabilidad (MIR 20, 177). El peor valor posible es 1 (los enfermos y los sanos tienen las mismas probabilidades de tener un resultado positivo, o bien un resultado negativo, lo que ocurriría si la S y E de la prueba son del 50%, esto es, la misma que el azar).

CAPACIDAD DIAGNÓSTICA	VALOR DE LA RPP	VALOR DE LA RPN
Suficiente	≥10	≤0,1
Moderada	5 - <10	>0,1 - 0,2
Escasa	2 - <5	>0,2 - 0,5
Insignificante	1 - <2	>0,5 - 1

Tabla 2. Capacidad diagnóstica de un test en función de la RPP y la RPN.

Parámetros de validez externa

La **validez externa** es la capacidad del test de generalizar los resultados obtenidos en la muestra a la **población** diana de la que se obtuvo la muestra. La validez interna es un requisito previo para la validez externa **(MIR)** *(si los resultados no son válidos para la muestra de sujetos, tampoco lo podrán ser para la población diana)*.

Valor predictivo positivo (VPP)

Capacidad del test de predecir si un sujeto que ha sacado positivo en el test va a estar realmente enfermo. Es la probabilidad de que un sujeto "+" (según el test) sea enfermo según el *gold standard*.

VPP = VP / total de positivos

Valor predictivo negativo (VPN) (MIR 23, 45; MIR)

Capacidad del test de predecir si un sujeto que ha sacado negativo en el test va a estar realmente sano. Es la probabilidad de que un sujeto "−" (según el test) sea sano según el *gold standard*.

VPN = VN / total de negativos

Valor global (VG)

Es la proporción de resultados verdaderos (verdaderos positivos y verdaderos negativos) del total de resultados de un test. Indica, por tanto, el porcentaje de veces que el test "acierta" en sus predicciones.

VG = (VP + VN) / n

Los parámetros de validez externa de un test diagnóstico dependen de la **probabilidad pre-test** de enfermedad de la población donde se aplique **(MIR 21, 50; MIR 15, 131)**. La probabilidad pre-test es la probabilidad que tiene un sujeto de tener una enfermedad antes de que se le realice un test diagnóstico. Depende de las características clínicas del sujeto (cuantos más síntomas y signos de la enfermedad, mayor probabilidad pre-test) y, fundamentalmente, es directamente proporcional a la **prevalencia** de enfermedad en la población **(MIR 13, 195; MIR 10, 195)**.

Así, si la prevalencia de una enfermedad es muy alta y un sujeto sale positivo en el test, será más probable que de verdad esté enfermo que si la prevalencia es muy baja. Por el contrario, si la prevalencia de enfermedad es baja y un sujeto sale negativo en el test, será más probable que esté de verdad sano **(MIR 12, 191; MIR 11, 191)**:

↑ prevalencia → ↑ VPP, ↓ VPN
↓ prevalencia → ↓ VPP, ↑ VPN

Antes hemos indicado que los tests muy sensibles son útiles cuando su resultado es negativo (descartan enfermedad), y los tests muy específicos cuando su resultado es positivo (confirman enfermedad). Esto es así por la relación entre la S y E con los valores predictivos de un test:

↑ S → ↓ TFN → ↑ VPN
↑ E → ↓ TFP → ↑ VPP

Recuerda...

Los valores **P**redictivos de un test **SÍ** dependen de la **P**revalencia, mientras que la S y E no dependen de la prevalencia.

Si una prueba diagnóstica tiene un **VPP 100%** significará que todos los individuos que den positivo en el test estarán enfermos, por ello se trata de una prueba **patognomónica (MIR 17, 129)**. Un hallazgo patognomónico se define como aquel signo o síntoma que demuestra de manera absoluta la existencia de una enfermedad.

5.2. Curvas ROC (de rendimiento diagnóstico) (MIR)

Cuando se define enfermedad o salud utilizando una **variable cuantitativa continua**, se debe definir un punto de corte a partir del cual consideramos que un sujeto es "positivo" y por tanto predecimos que estará enfermo.

Ejemplo: se considera diabético a un individuo que tenga ≥126 mg/dl de glucemia en ayunas en al menos dos determinaciones separadas en el tiempo.

En las variables cuantitativas, a medida que llevamos el punto de corte que define enfermedad a niveles más "enfermos", seremos más específicos pero menos sensibles **(MIR 14, 206; MIR 12, 193)**. Por el contrario, si llevamos el punto de corte a niveles más "sanos", seremos más sensibles y menos específicos. Así, podemos afirmar que para las variables cuantitativas la S y la E son inversamente proporcionales: **al aumentar la S disminuye la E, y viceversa**.

Punto de corte más "**E**nfermo" → ↑ **E** y ↓ S
Punto de corte más "**S**ano" → ↑ **S** y ↓ E

Ejemplo: si en lugar de utilizar un nivel de glucemia de 126 mg/dl para definir diabetes, llevamos el punto de corte a un nivel más "enfermo" (p. ej., a 150 mg/dl), el nuevo punto de corte será más específico (habrá menos número de FP, ya que casi todos los pacientes con glucemia >150 mg/dl serán de verdad diabéticos –VP–) pero menos sensible (habrá más número de FN, ya que muchos pacientes diabéticos tienen glucemias menores a 150 mg/dl y no vamos a ser capaces de diagnosticarlos).

Las **curvas ROC** muestran el nivel de S y de E que obtenemos con cada posible punto de corte de la variable cuantitativa, lo que nos permite escoger el mejor punto de corte (aquel con una mejor relación entre sensibilidad y especificidad). Gráficamente se representan poniendo la S en el eje de ordenadas, y la TFP (1 – E) en el eje de abscisas. El mejor punto de corte es aquel que corta la bisectriz de la curva ROC.

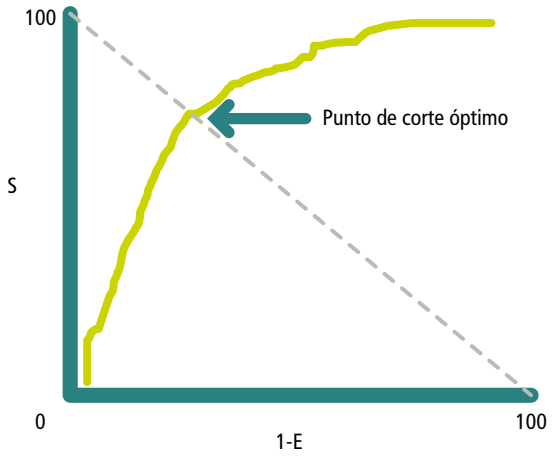


Figura 1. Curva ROC.

El **área bajo la curva** de las curvas ROC representa el grado de **validez global** del test **(MIR)**. Cuando comparamos varios tests diagnósticos, será mejor aquel cuya área bajo la curva ROC sea mayor (el vértice de la curva estará situado más cerca del ángulo superior izquierdo).

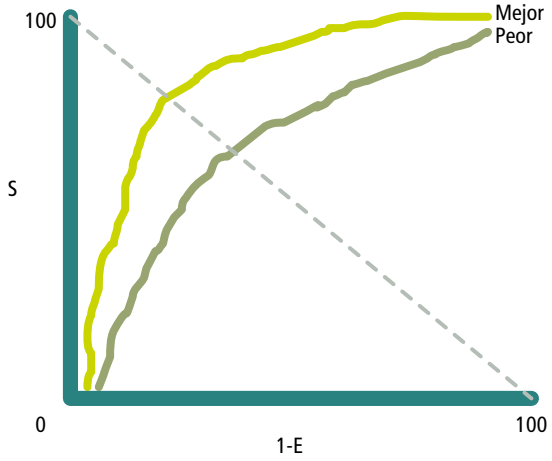


Figura 2. Comparación de la validez global de varios tests mediante sus curvas ROC.

5.3. Test de *screening* y test de confirmación

Test de *screening* (MIR 13, 197)

En muchas ocasiones, el proceso diagnóstico de una enfermedad se inicia realizando un **test de *screening*** (o cribado poblacional). El objetivo de un test de *screening* es doble: detectar casos precoces (presintomáticos) de enfermedad, y de manera más importante descartar a los sujetos sanos (que sacan negativo en el test). Por tanto, los tests de *screening* deben ser **muy sensibles (MIR 18, 222; MIR)**; los pacientes que den negativo en el test sabremos que están sanos, y a los pacientes que den positivo se aplicará luego un test más específico para confirmar la enfermedad.

La característica **más importante** de los tests de *screening* es que deben tener un **alto VPP** en la población donde se apliquen **(MIR)**. Si un test de *screening* se aplica en una población de muy baja prevalencia de enfermedad, la mayoría de sujetos que den positivo en el test serán realmente FP; nos veremos obligados a realizar en balde muchos tests diagnósticos de confirmación, lo cual supondrá un coste económico inasumible.

Así, no todas las enfermedades son susceptibles de *screening*, sino que se deben cumplir una serie de requisitos para que éste se pueda instaurar:

Criterios de la enfermedad

- Enfermedad frecuente en la población estudiada.
- Enfermedad grave que no debe pasar desapercibida (si no se diagnostica a tiempo empeora el pronóstico).
- La fase presintomática no debe ser corta **(MIR)**.
- Se debe conocer la historia natural de la enfermedad.
- La enfermedad debe tener un tratamiento más eficaz si se aplica en fase presintomática que si se aplica en fase sintomática.

Criterios del test

- Fácil de realizar.
- Inocuo.
- De coste razonable (pero no tiene por qué ser menos costoso que tratar un caso de la enfermedad) **(MIR)**.
- Buenos valores de validez (primando la S sobre la E) y reproducibilidad.
- Aceptable y visto como necesario por la comunidad.

Test de confirmación

Los tests que se utilizan para confirmar la presencia de enfermedad deben ser **muy específicos (MIR 14, 207)** (para que los sujetos positivos tengan muchas probabilidades de ser realmente enfermos).

Las principales circunstancias en las que es importante utilizar tests de confirmación para diagnosticar de forma definitiva una enfermedad son:

- Enfermedades graves pero sin tratamiento eficaz.
- Los falsos positivos pueden suponer un trauma emocional **(MIR)**.
- Tratar los falsos positivos puede tener graves consecuencias.
- Enfermedades de prevalencia muy baja **(MIR)**.

Tema 6

Medidas en epidemiología

Autores: *Víctor Rodríguez Domínguez (2), Rafael Leal Hidalgo (47), Carlos Corrales Benítez (16), Eduardo Franco Díez (5), Héctor Manjón Rubio (42).*

ENFOQUE MIR

Tercer tema en importancia del manual. Casi todos los años caen preguntas en las que piden interpretar el intervalo de confianza de una medida de asociación (habitualmente del RR) o de impacto. Además, los conceptos teóricos también son preguntados, y pueden caer **problemas** (siendo el más frecuente el cálculo del NNT).

6.1. Medidas de frecuencia de una enfermedad

Prevalencia (MIR)

Es la proporción de individuos de una población que padecen una determinada enfermedad **en un momento dado** (MIR).

Es muy útil para valorar la extensión de **enfermedades crónicas**. Sin embargo, como sólo se evalúa un momento concreto y no un periodo de tiempo, no es útil para el estudio enfermedades agudas *(las enfermedades agudas aparecen y desaparecen, de modo que al estudiar un momento del tiempo concreto es probable no encontrar la enfermedad)*.

Si se desea estimar la prevalencia de una enfermedad, lo más eficiente es diseñar para ello un **estudio transversal**. Sin embargo, en los estudios longitudinales se podría también determinar la prevalencia en cualquier momento dado.

La prevalencia de una enfermedad **aumenta en las siguientes circunstancias**:

- Aumento de la incidencia de la enfermedad (aumento de casos nuevos).
- Aumento de duración de la enfermedad (si disminuye su mortalidad).
- Descenso de la tasa de curación de la enfermedad.
- Mejora de los métodos diagnósticos de una enfermedad (se descubrirán más casos).
- Inmigración de casos enfermos o emigración de sujetos sanos.

Incidencia (incidencia acumulada) (IA)

Es la proporción de **casos nuevos** de una enfermedad que aparecen en una población en un determinado **periodo de tiempo**, con respecto al total de la población que es **susceptible** de enfermar (MIR 20, 31; MIR). *Por ejemplo, si se desea medir la incidencia de una enfermedad crónica, habrá que calcular la proporción de casos nuevos respecto de los sujetos que no tengan dicha enfermedad (los que ya la tienen no son susceptibles de volver a enfermar).*

Indica **la probabilidad que tiene un sujeto sano de enfermar** a lo largo del periodo de tiempo que se haya tenido en cuenta para el cálculo de la incidencia **(riesgo individual de enfermar)** (MIR).

Para calcularla son necesarios estudios longitudinales prospectivos, ya que necesitamos un periodo de seguimiento que vaya hacia el futuro para cuantificar los casos nuevos que van apareciendo.

Densidad de incidencia (DI)

(MIR 20, 37; MIR 14, 196; MIR 10, 194)

Es la **velocidad** con la que se propaga una enfermedad en una población, e indica el número de casos nuevos que aparecen por **unidad de tiempo**. El tiempo que se utiliza como unidad de medida es la suma del tiempo que ha estado expuesto a la enfermedad cada individuo hasta que la contrae: **suma de los tiempos de observación**.

En el momento que un individuo enferma, si ya no puede volver a enfermar finaliza su tiempo de observación. Si un individuo no enferma a lo largo de todo el periodo de seguimiento, su tiempo de observación será lo que dure dicho periodo.

Para calcularla también son necesarios **estudios longitudinales prospectivos**.

$$\textbf{Prevalencia} = \frac{\text{n.º de casos en un momento puntual}}{\text{población}}$$

$$\textbf{IA} = \frac{\text{n.º de casos nuevos a lo largo de un periodo de tiempo}}{\text{población susceptible de enfermar al inicio del periodo}}$$

$$\textbf{DI} = \frac{\text{n.º de casos nuevos a lo largo de un periodo de tiempo}}{\sum \text{t de observación de cada individuo susceptible de enfermar}}$$

Tabla 1. Medidas de frecuencia de una enfermedad.

6.2. Medidas de fuerza de asociación (medidas de efecto)

Todas ellas son razones que se calculan mediante el **cociente** entre el riesgo que presentan los sujetos expuestos a un determinado factor (de riesgo o protector), y el riesgo que presentan los no expuestos.

Así, miden **cuántas veces** es más frecuente la enfermedad en el grupo expuesto respecto al no expuesto **(MIR)**. Miden pues la "**fuerza de asociación**" entre un factor causal y su efecto **(MIR)**.

Su resultado oscila entre 0 e infinito (rango) **(MIR 20, 34)**, y **no tienen unidades (MIR)**:

- Si el resultado es <1: el factor estudiado es un **factor protector**.
- Si el resultado es >1: el factor estudiado es un **factor de riesgo**.
- Si el resultado es = 1: no existe relación causal entre el factor y la enfermedad (no es factor de riesgo ni de protección).

Cuando se extrapola el resultado obtenido a una **población** a partir de una muestra (estadística inferencial), el intervalo de confianza (IC) obtenido informa sobre la significación estadística del resultado. El valor de la medida de asociación no tiene por qué estar en el centro del IC **(MIR 11, 174)**, ya que puede que existan más probabilidades de que el riesgo sea mayor o menor que ese valor que hemos obtenido, que viceversa.

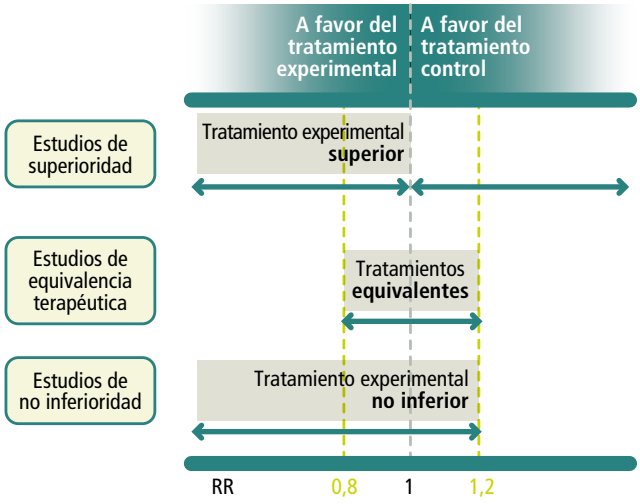


Figura 1. Interpretación del IC de las medidas de fuerza de asociación.

Diseño de superioridad

(MIR 15, 199; MIR 14, 205; MIR 12, 175; MIR 12, 182)

- Si el IC incluye el **1**, no es estadísticamente significativo.
- Si el IC no incluye el **1**, es estadísticamente significativo.

*En el caso de que el riesgo del grupo expuesto y no expuesto fuera el mismo, el cociente entre esos riesgos sería = 1, así que si el "1" está incluido en el intervalo de confianza, significará que existen probabilidades de que el factor estudiado sea tanto un factor protector ("parte" del intervalo de confianza que sea <1) como un factor de riesgo ("parte" del intervalo de confianza que sea >1). El 1 es el "**valor de no significación**".*

Diseño de no inferioridad

Para poder establecer que una intervención es **no inferior** a otra, el IC para la intervención experimental debe encontrarse totalmente **por debajo de 1,2 (MIR 11, 187)** (menos de un 20% de riesgo adicional respecto al fármaco control), si el límite de no inferioridad (delta) se establece en el 20%.

Diseño de equivalencia terapéutica

Para poder establecer que dos tratamientos son **equivalentes terapéuticos** entre sí, el IC de cualquiera respecto al otro debe encontrarse delimitado **entre 0,8 y 1,2 (MIR)** (si hemos establecido unos límites del 20%)

Según qué medida de frecuencia de la enfermedad (del "riesgo" de enfermar) estemos utilizando, usaremos una u otra medida de asociación:

Riesgo relativo (RR)

Es la medida que se utiliza cuando disponemos de **incidencias acumuladas**. Como requiere del cálculo de incidencias, sólo se podrá calcular en estudios que presenten un **seguimiento prospectivo**: estudio de cohortes, ensayo clínico, etc. Es la medida de efecto que **mejor estima** el riesgo real.

Odds ratio o razón de desventaja (OR)

Es la medida que se utiliza en los estudios con un **seguimiento retrospectivo** (estudio de casos y controles), en los cuales nopodemos calcular incidencias, sino las **prevalencias** del factor de riesgo en el grupo enfermo y en el grupo sano.

Es peor estimador del riesgo real que el riesgo relativo y tiende a **sobreestimar la fuerza de asociación**. Para que su valor estime bien el RR, los controles y los casos deben provenir de la misma población, y la **incidencia de la enfermedad debe ser <10%** (esto es, en enfermedades poco frecuentes, se aproxima bastante al RR) **(MIR)**.

Razón de prevalencia o de proporciones (RP)

Es la medida que se utiliza en los estudios **sin seguimiento** (estudios transversales, etc.), en los cuales lo único que podemos calcular es la prevalencia en un momento puntual de la enfermedad en el grupo de expuestos y en el de sanos.

Su cálculo matemático es idéntico al del RR, pero **es el peor estimador del riesgo real** por el diseño de los estudios a partir de los que se calcula (que no tienen seguimiento, por lo que nunca pueden demostrar causalidad).

(Ver tabla 2)

	ENFERMOS	SANOS
FACTOR PRESENTE	a	b
FACTOR AUSENTE	c	d

$$\textbf{RR} = \frac{\text{IA expuestos (Ie)}}{\text{IA en no expuestos (Io)}} = \frac{a/(a+b)}{c/(c+d)}$$

$$\textbf{OR} = \frac{\text{"Odds" del grupo enfermo}}{\text{"Odds" del grupo sano}} = \frac{\dfrac{\text{prevalencia del factor en enfermos}}{\text{prevalencia de no tener factor en enfermos}}}{\dfrac{\text{prevalencia del factor en sanos}}{\text{prevalencia de no tener factor en sanos}}} = \frac{\dfrac{a/(a+c)}{c/(a+c)}}{\dfrac{b/(b+d)}{d/(b+d)}} = \frac{a \cdot d}{b \cdot c}$$

$$\textbf{RP} = \frac{\text{prevalencia de enfermedad en expuestos}}{\text{prevalencia de enfermedad en no expuestos}} = \frac{a/(a+b)}{c/(c+d)}$$

Tabla 2. Medidas de fuerza de asociación **(MIR 15, 181; MIR 15, 189; MIR 13, 181)**.

6.3. Criterios de causalidad de Bradford Hill

El hecho de que exista una determinada fuerza de asociación entre un factor y una enfermedad NO implica necesariamente que dicho factor sea un factor causal de dicha enfermedad.

Para que se establezca una **relación de causalidad** se deben cumplir varios de los siguientes criterios (no hace falta que se cumplan todos):

Criterios de validez interna

- **Secuencia temporal**: la causa debe preceder al efecto. Es el único criterio de causalidad **imprescindible** **(MIR 20, 32; MIR)**.
- **Fuerza de asociación** **(MIR)**: a mayor magnitud de la medida de fuerza de asociación, mayor es la probabilidad de que exista una relación causal.
- **Efecto dosis-respuesta (gradiente biológico)**: a mayor dosis o tiempo de exposición al factor causal, mayor es el riesgo de enfermar.

Criterios de coherencia científica

- **Consistencia**: los resultados de un estudio que sugiera causalidad deben ser reproducibles por otros investigadores y arrojar resultados similares.
- **Coherencia**: los resultados de los estudios que traten de establecer la relación causal entre un factor y un efecto deben ser similares entre sí.

Figura 2. Sir Austin Bradford Hill (1897-1991).

- **Plausibilidad biológica**: existencia de un mecanismo fisiopatológico conocido que explique la posible relación causal.
- **Especificidad de asociación**: si la posible causa conduce a un único efecto, y viceversa, el efecto sólo parece estar causado por un único factor, la verosimilitud de la relación causal aumenta.

- **Analogía**: factores causales similares al estudiado producen efectos similares.
- **Demostración experimental**: existencia de asociación entre el factor y el efecto en estudios experimentales. Es el criterio de causalidad **más potente**.

No son criterios de causalidad (MIR)

- La existencia de asociación estadísticamente significativa (p <0,05).
- La respuesta a un tratamiento concreto.

6.4. Medidas de impacto

Cuantifican cuál es el impacto de una medida preventiva al aplicarla en una población (al suprimir un determinado factor de riesgo, o al implementar un determinado factor protector). Utilizan **incidencias acumuladas**, por lo que se calculan **sólo en estudios con seguimiento prospectivo** (estudio de cohortes, ensayos clínicos, etc.).

En el cálculo de todas ellas existe una resta entre la incidencia en expuestos y no expuestos. Por tanto, el "**valor de no significación**" (aquel que indica que el riesgo en expuestos y no expuestos es el mismo), es el "**0**". En estadística inferencial, y para los estudios de superioridad **(MIR)**:

- Si el IC incluye el 0, no es estadísticamente significativo.
- Si el IC no incluye el 0, es estadísticamente significativo.

Medidas de impacto absolutas

Indican el beneficio absoluto (número de casos evitados por cada 100 personas en riesgo) que se obtiene al retirar un factor de riesgo o implementar un factor protector totalmente en una muestra o población en riesgo. Por lo tanto, son medidas útiles en **Salud Pública** (al permitirnos calcular, si conocemos el tamaño de una población, el número total de casos que evitaríamos en ella).

Riesgo atribuible, exceso de riesgo o diferencia de incidencias (RA, ER)
(MIR 17, 134; MIR 15, 181)

Medida de impacto absoluta utilizada para **factores de riesgo**. Indica el exceso de riesgo asociado a la exposición, y que podría evitarse si se eliminara ésta (número de casos evitados por cada 100 pacientes con el factor de riesgo a los que les quitas dicho factor).

$$RA = Ie - Io$$

Ejemplo: RA = 6% significa que por cada 100 expuestos hay seis casos más de enfermedad que por cada 100 no expuestos. Si eliminásemos el factor de riesgo en un grupo de expuestos, evitaríamos por cada 100 expuestos esos seis casos de más.

Reducción absoluta de riesgo (RAR)
(MIR 14, 195; MIR 11, 186)

Medida de impacto absoluta utilizada para **factores de protección**. Indica la reducción en la incidencia de enfermedad que conseguiríamos al implementar un factor protector en un grupo en riesgo (número de casos evitados por cada 100 pacientes no protegidos, a los que se les proporciona el factor protector).

$$RAR = Io - Ie$$

Ejemplo: RAR = 5% significa que por cada 100 personas no protegidas hay cinco casos más de enfermedad que por cada 100 personas con el factor de protección. Si a las personas no protegidas les proporcionáramos el factor de protección, evitaríamos esos cinco casos que tienen de más.

Número necesario de pacientes a tratar (NNT)
(MIR 22, 46; MIR 21, 54; MIR 19, 119; MIR 17, 118; MIR 16, 191; MIR 13, 182; MIR 11, 184)

Es el número de pacientes que se debe tratar con un factor protector para prevenir un evento. Se obtiene a partir del inverso de la RAR:

$$NNT = 100 / RAR \text{ (expresando el RAR en \%)}$$

Se debe redondear al entero superior. Al igual que para el resto de medidas de impacto, el "0" contenido en el intervalo de confianza indica no significación estadística.

Al igual que podemos calcular el número de pacientes que hay que tratar con un factor protector para prevenir un evento, usando para ello la RAR, también podemos calcular el número de pacientes que hay que "dañar" con un factor de riesgo para provocar un caso de enfermedad (**NNH: número necesario de pacientes a dañar**) **(MIR 15, 198)**, usando para ello el RA:

$$NNH = 100 / RA \text{ (expresando el RA en \%)}$$

Medidas de impacto relativas

Indican el beneficio relativo (porcentaje de casos evitados del total de casos que padece una población en riesgo) que se obtiene al retirar un factor de riesgo o implementar un factor protector en una población en riesgo. Por lo tanto, son medidas útiles en **Epidemiología Clínica** (al sernos un indicador del porcentaje de riesgo de enfermar que evitamos en cada sujeto en riesgo).

Fracción atribuible o fracción etiológica de riesgo (FA, FER)

Medida de impacto relativa utilizada para **factores de riesgo**. Es la proporción de casos nuevos entre los expuestos que es atribuible a la exposición.

$$FA = (Ie - Io) / Ie$$

Ejemplo: FA = 40% significa que de cada 100 casos de enfermedad que aparecen en un grupo de expuestos, 40 se deben a esa exposición (60 se deberán a otras causas). Así, si un individuo expuesto elimina su factor de riesgo (p. ej., deja de fumar) su riesgo de enfermar disminuirá un 40%.

Reducción relativa de riesgo (RRR) (MIR)

Medida de impacto relativa utilizada para **factores de protección**. Es la proporción de casos nuevos, entre los sujetos que no tienen el factor protector, que es atribuible a la ausencia de la protección que dicho factor confiere.

$$RRR = (Io - Ie) / Io = 1 - RR$$

Ejemplo: RRR = 35% significa que de cada 100 casos de enfermedad que aparecen en un grupo sin el factor protector (p. ej., no vacunados), 35 se deben a no tener el factor protector. Así, si un individuo adquiere el factor protector (p. ej., se vacuna) su riesgo de enfermar disminuirá un 35%.

Tema 7

Tipos de estudios epidemiológicos

Autores: *Rafael Leal Hidalgo (47), Eduardo Franco Díez (5), Héctor Manjón Rubio (42), Julio Sesma Romero (36).*

ENFOQUE MIR

Tema más importante de la asignatura y de gran importancia para el MIR en general. Todos los años caen varias preguntas teóricas para elegir el tipo de estudio epidemiológico que se presenta en el enunciado, así como sobre las diferencias entre el estudio de casos y controles y el estudio de cohortes y sobre el ensayo clínico. En los últimos años han ganado importancia temas relacionados con Medicina Basada en la Evidencia, como los grados de recomendación o el metaanálisis y sus características.

En función de sus **objetivos**, existen dos grandes grupos de estudios epidemiológicos:

- **Estudios descriptivos** (MIR 18, 213): su objetivo es **describir** la naturaleza y magnitud de un problema de salud, entre quiénes y dónde se produce y otras características similares.

  Se consideran descriptivos los siguientes estudios epidemiológicos:

  - Comunicación de un caso/serie de casos: siempre **descriptivos**.
  - Estudio transversal y estudio ecológico: pueden ser también **analíticos**, pero en dicho caso no pueden demostrar hipótesis.

- **Estudios analíticos**: su objetivo es establecer la **relación** entre una determinada exposición y la aparición de un determinado problema de salud.

  Se consideran analíticos los siguientes estudios epidemiológicos:

  - Casos y controles.
  - Cohortes (MIR 10, 185).
  - Estudios experimentales.
  - Estudios cuasi-experimentales.

7.1. Estudios observacionales (MIR 15, 197)

Se distinguen de los estudios experimentales en la **ausencia de intervención** por parte del investigador, que se limita a observar lo que ocurre en la **práctica clínica habitual** (MIR).

Existen distintos estudios observacionales que se diferencian por el **tipo de seguimiento** realizado a los pacientes.

- **Sin seguimiento (estudios transversales)**: estudio transversal y estudio ecológico.
- **Con seguimiento (estudios longitudinales)**:
  - Retrospectivo: estudio de casos y controles.
  - Prospectivo: estudio de cohortes; estudio de tendencias temporales.

Comunicación de un caso/serie de casos

Son estudios que describen las características o el manejo clínico realizado en un paciente o grupo de pacientes con un diagnóstico similar.

Generan **nuevas hipótesis** de trabajo, pero no permiten confirmar hipótesis ya que **carecen de un grupo control** (principal limitación).

Estudio transversal (estudio de prevalencia, estudio de corte)

(MIR 17, 131; MIR 16, 193; MIR 13, 178)

Es un estudio observacional de **base individual** (la unidad del estudio es el individuo) que **no presenta seguimiento** de los pacientes, esto es, que sólo estudia las características que tienen los pacientes en el **presente**: trata de describir o estudiar relaciones causales entre exposiciones y problemas de salud presentes en **un momento puntual**.

Utiliza como medida de fuerza de asociación la **razón de prevalencias**.

Ventajas

- Rápido, barato y reproducible, al prescindir del seguimiento de los pacientes.
- Tipo de diseño adecuado para evaluar la **validez de una prueba diagnóstica**.
- Tipo de diseño más eficiente para estimar la **prevalencia** de una enfermedad **(MIR 10, 182; MIR)** *(cualquier estudio puede medir en un momento puntual la prevalencia de enfermedad, pero el estudio transversal es el más barato)*.
- Útil para el estudio de **enfermedades crónicas**.
- Útil para **planificación sanitaria (Salud Pública)**, ya que permite de forma barata el estudio de enfermedades crónicas, que son las que más recursos sanitarios consumen.

Inconvenientes

- Muy sensible a los **sesgos**.
- No permite valorar la **secuencia temporal** (ya que estudia la presencia de la exposición y la enfermedad en el mismo momento). Por lo tanto, **no permite demostrar hipótesis etiológicas (causalidad)**, sino que sólo las **genera** (MIR).
- No es útil para el estudio de **enfermedades agudas** (ya que no permite medir incidencias por no tener seguimiento) ni raras.

Estudio ecológico (estudio de correlación ecológica o de riesgo agregado)

(MIR 22, 41; MIR 16, 192; MIR 13, 183; MIR 11, 181)

Es un estudio idéntico al estudio transversal, con la única diferencia de tener una **base comunitaria** en lugar de tener una base individual.

Por lo tanto, es útil para generar hipótesis de trabajo pero no las demuestra (MIR).

Características de una **base comunitaria**:

- Utiliza datos recogidos de **grupos de personas** (en lugar de individuos), formados en general por criterios geográficos (países, comunidades autónomas, ciudades...).
- Para recopilar los datos, se acude a **registros**, que son quienes proporcionan los datos (esto es, otra persona ha recogido los datos individuales previamente y ha extrapolado los datos poblacionales, que son los que utilizaremos).
- Utilizan **datos indirectos o secundarios** (MIR) (recogidos por otras personas). Los datos indirectos son de **peor calidad** (más riesgo de **sesgos**) que los directos ya que no podemos controlar los criterios o instrumentos de medida utilizados por la persona que los recogió.
- Los datos que recogemos son los **promedios** (se habla de promedios en lugar de medias o de porcentajes) de la característica estudiada en cada uno de los grupos de personas estudiado.

Estudio de casos y controles

(MIR 18, 224; MIR 15, 183; MIR 14, 201; MIR 12, 180; MIR 12, 183; MIR 12, 184; MIR 10, 183)

Es un estudio observacional de base individual y con **seguimiento retrospectivo** (desde el presente hacia el pasado). Dicho seguimiento no es un seguimiento real que implique ver a los pacientes varias veces (un estudio de casos y controles sólo requiere ver a cada paciente una única vez), sino que es un seguimiento "virtual" que se realiza utilizando la **memoria** del paciente para que nos cuente datos de su pasado.

*El seguimiento retrospectivo, por lo tanto, está sujeto a las limitaciones de la memoria humana y es peor (más sujeto a **sesgos**) que el seguimiento prospectivo.*

En un estudio de casos y controles se selecciona un grupo de sujetos **enfermos (casos)** y un grupo de **sujetos sanos (controles)** respecto al problema de salud que se quiere analizar, y se evalúa la proporción de sujetos de cada uno de los grupos que estaba expuesta en el pasado a un supuesto **factor de riesgo o protector**. Pueden escogerse varios controles para cada caso (MIR).

Para evitar sesgos de memoria es importante ser igual de exhaustivo en la anamnesis de los casos y de los controles (MIR).

- Utiliza como medida de fuerza de asociación la **odds ratio**, que **sobreestima** el riesgo con respecto al RR salvo en el estudio de enfermedades raras.
- Sus ventajas e inconvenientes están recogidos en la **tabla 1**.

Estudio de casos y controles anidado en una cohorte

(MIR 19, 120; MIR 17, 132; MIR 13, 184; MIR 11, 180)

Es un caso particular del estudio de casos y controles que se utiliza cuando en el momento actual **no existe un número de casos suficiente** para realizar el estudio. Esto suele ocurrir cuando se estudian enfermedades agudas y epidémicas, en las que los casos van apareciendo poco a poco a lo largo de un periodo de tiempo.

Este estudio consiste en realizar un seguimiento prospectivo de una cohorte de sujetos (de la población diana de nuestro estudio), e ir seleccionando los casos a medida que aparecen hasta alcanzar el tamaño muestral deseado. Cada vez que aparece un caso, se seleccionan de manera aleatoria entre los sujetos sanos de la cohorte el número de controles que se haya previsto para cada caso, y en ese momento **se recogen los datos de forma retrospectiva** (se pregunta al caso y a los controles por su pasado).

Así, sus características son idénticas a las del estudio de casos y controles convencional, y sólo se diferencia en que los pacientes se reclutan poco a poco en lugar de todos a la vez.

Los estudios de casos y controles realizados utilizando los datos de los **registros poblacionales** de enfermedades suelen considerarse anidados, dado que los registros poblacionales suelen elaborarse de forma prospectiva (MIR 14, 199).

Estudio de cohortes

(MIR 23, 43; MIR 22, 48; MIR 21, 45; MIR 20, 30; MIR 19, 121; MIR 18, 212; MIR 15, 182; MIR 13, 179; MIR 12, 179; MIR 11, 179; MIR 10, 180)

Es un estudio observacional de base individual y con **seguimiento prospectivo** (desde el presente hacia el futuro). Dicho seguimiento es real, y consiste en ver al paciente hoy y volverle a ver en sucesivas ocasiones en el futuro hasta que finalice el periodo de seguimiento. Es el **mejor estudio observacional** para demostrar hipótesis.

En un estudio de cohortes se sigue prospectivamente a **dos grupos de individuos sanos** con respecto al problema de salud que se quiere analizar: un grupo que está **expuesto** a un factor de riesgo o protector, y un grupo **no expuesto** (MIR). Se analiza la incidencia de enfermedad que aparece en cada uno de esos dos grupos a lo largo del periodo de seguimiento (MIR).

- Utiliza como medida de fuerza de asociación el **RR**.
- Sus ventajas e inconvenientes están recogidos en la **tabla 1**.

CASOS Y CONTROLES	COHORTES
Seguimiento retrospectivo	Seguimiento prospectivo
Calculan prevalencia de exposición	Calculan incidencias de enfermedad
OR: sobreestima la fuerza de asociación, salvo en enfermedades raras	**RR**: mejor estimador de la asociación
Baratos, rápidos, reproducibles	Caros, lentos, poco reproducibles
Peores para demostrar hipótesis	Mejores para demostrar hipótesis
Más sensible a sesgos **(MIR)**	Menos sensible a sesgos
Útiles para estudiar **enfermedades raras** o con largo periodo de latencia **(MIR 17, 120; MIR 17, 133; MIR 11, 182)**	Útiles para estudiar **exposiciones raras** y para enfermedades agudas
Permiten estudiar **multicausalidad** (varias causas de una enfermedad)	Permiten estudiar **multiefectividad** (varios efectos de la misma exposición)
Dificultad: establecer el grupo control*	**Dificultad**: sensible a pérdidas**

*La principal dificultad de los estudios de **casos y controles** es seleccionar el **grupo control**, ya que óptimamente debe tener las mismas características que el grupo de casos para minimizar el riesgo de aparición de sesgos de selección o de sesgos por factor de confusión. Para conseguirlo se emplean técnicas de **apareamiento** **(ver tema 8.2. Errores sistemáticos (sesgos))**.

Los estudios de **cohortes (y cualquier estudio prospectivo) son sensibles a la aparición de pérdidas, ya que requieren ver a los pacientes en el futuro y eso puede no ser posible (fallecimientos, cambio de domicilio, incomparecencia del paciente...). En cambio, los estudios de **casos y controles** sólo requieren ver a cada paciente una vez, por lo que **no pueden tener pérdidas**.

Tabla 1. Diferencias entre los estudios de casos y controles y los estudios de cohortes **(MIR)**.

Estudio de cohortes históricas (cohortes retrospectivo)

(MIR 18, 211; MIR 13, 186)

Es un estudio de cohortes en el que el seguimiento de los pacientes se realiza **desde el pasado hacia el presente**, en lugar de hacerlo desde el presente hacia el futuro. Fijaos que la dirección del seguimiento sigue siendo **prospectiva**.

*El nombre "estudio de cohortes retrospectivo" es por lo tanto desafortunado ya que el seguimiento **no es retrospectivo** sino prospectivo. Es mejor utilizar el término estudio de cohortes históricas, aunque los dos términos son correctos.*

El seguimiento se realiza de modo **indirecto** a través de la historia clínica del paciente, comenzando por la primera página (pasado) y acabando en la última (presente).

Sus características son idénticas a las del estudio de cohortes convencional, excepto en los siguientes aspectos:

Ventaja: al no requerir de un seguimiento real de pacientes, es más barato, rápido y reproducible.

Inconveniente: utiliza **datos indirectos** (recogidos en la historia por terceras personas), por lo que es más sensible a los sesgos.

(Ver figura 1)

Estudio de series o tendencias temporales

Se puede definir como un estudio de cohortes pero que tiene una **base comunitaria** (aplicándose por tanto las características de la base comunitaria). Así, es también similar al estudio ecológico, pero con un seguimiento prospectivo.

Son grandes registros con seguimientos prospectivos largos en los que se trata de establecer cómo evoluciona un problema de salud en una o varias poblaciones a lo largo del tiempo **(MIR)**.

En el MIR nos lo suelen mostrar relacionado a la evolución temporal de problemas medioambientales más que de enfermedades.

Ejemplos: evolución temporal de los niveles de contaminación de un río o de un mar.

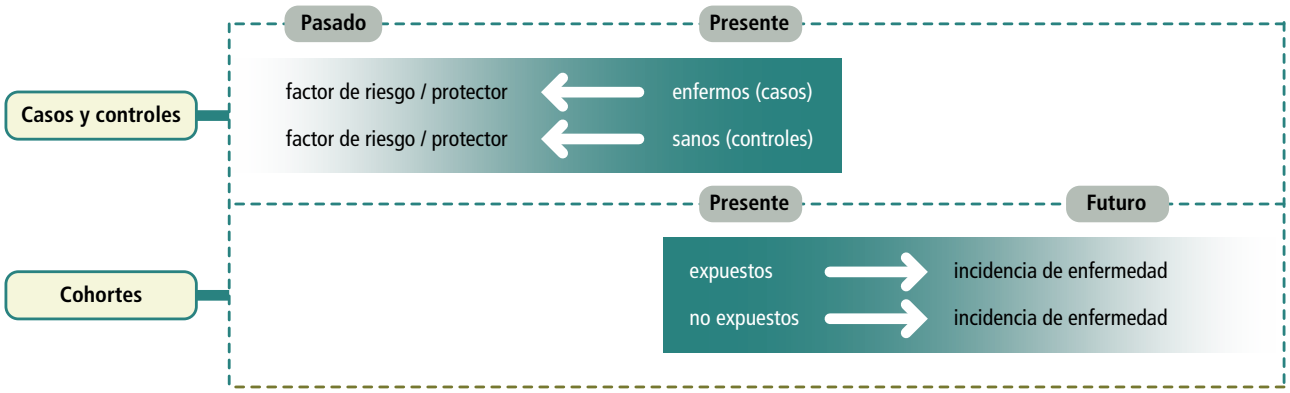


Figura 1. Estudio de casos y controles y estudio de cohortes.

7.2. Estudios experimentales

Se distinguen de los estudios observacionales en la **presencia de intervención** por parte del investigador (MIR 14, 193): el investigador introduce de forma directa una nueva medida diagnóstica, terapéutica o preventiva en una determinada muestra de individuos.

Son los **mejores estudios para demostrar hipótesis** (MIR 10, 181; MIR), y son **menos sensibles a los sesgos** (MIR 13, 180) que los estudios observacionales. Por el contrario, todos requieren un seguimiento prospectivo y son caros, lentos y poco reproducibles. Además, están sujetos a **problemas éticos** (ya que se interviene activamente sobre la salud de las personas).

Los tipos de estudios experimentales que se diferencian por el modo de **asignación** de la intervención:

Estudios "experimentales"

La asignación de intervención o no intervención a cada individuo se realiza de manera **aleatoria**, de modo que **es el azar el que forma los distintos grupos** (grupo intervención/grupo no intervención) que se van a comparar entre sí (MIR 23, 52).

Se consideran mejores estudios para demostrar hipótesis que los cuasi-experimentales.

Para poder llevar a cabo cualquier estudio experimental, es necesario obtener autorización por parte de la **Agencia Española del Medicamento** (MIR 10, 191), así como un dictamen favorable del **Comité Ético de Investigación Clínica (CEIC)** de cada uno de los Centros médicos que participen en el estudio (MIR 12, 187).

- **Ensayo de campo** (MIR): estudio experimental cuyo objetivo es la prevención de una enfermedad mediante la aplicación de una medida preventiva (p. ej., una vacuna) a un grupo de **sujetos sanos**, cuyos resultados se compararán con los de otro grupo de sujetos sanos a los que no se aplicó la medida preventiva.
- **Ensayo clínico** (MIR 15, 196; MIR 12, 178; MIR 10, 184): estudio experimental cuyo objetivo es el **tratamiento** de una enfermedad mediante la aplicación de una intervención (p. ej., un fármaco) a un grupo de **sujetos enfermos**, cuyos resultados se compararán con los de otro grupo de enfermos a los que no se aplicó el tratamiento.

  La declaración **CONSORT** (*Consolidated Standards of Reporting Trials*) (MIR 11, 183) tiene como objetivo mejorar la redacción de artículos sobre ensayos clínicos aleatorizados. Establece una serie de normas básicas de redacción que los autores deben cumplir, y establece una lista de puntos que los lectores deben chequear para comprobar que los datos importantes están incluidos en el artículo, favoreciendo así la comprensión y capacidad crítica de los lectores.

  La declaración STROBE (*Strengthening the Reporting of Observational studies in Epidemiology*) tiene un objetivo similar a la CONSORT, pero para estudios observacionales (transversales, casos y controles y cohortes). Para los metaanálisis, existe la declaración PRISMA.

Estudios "cuasi-experimentales"

La asignación de la intervención se realiza por un **método no aleatorio**: los grupos que van a participar en el estudio no se forman mediante el azar. Si **sólo hay un grupo** de pacientes en un estudio con intervención, siempre será cuasi-experimental, ya que la intervención se realizará en dicho grupo y el azar no podrá decidir nada.

*Por ejemplo, si los grupos que van a participar ya existen de manera natural (p. ej., dos localidades) y se elige cuál de los dos grupos recibirá la intervención mediante el azar (tirando una moneda al aire) no se está aleatorizando, ya que la aleatorización implica **formar los grupos** mediante el azar.*

- **Estudio de intervención comunitaria** (MIR): son estudios con **base comunitaria**, y en general para valorar medidas preventivas. Casi siempre son cuasi-experimentales (pero podrían ser aleatorizados) ya que, al utilizar una base comunitaria, los grupos de estudio suelen estar preformados y en esos casos es imposible formarlos mediante el azar.

  *Los estudios de intervención comunitaria son habitualmente la alternativa a un **ensayo de campo** cuando es muy difícil utilizar una base individual. Por ejemplo, si se realiza un estudio para demostrar si el consumo de agua fluorada previene la aparición de caries, aleatorizar individuos y que unos tomen agua fluorada y otros agua normal puede ser muy difícil. En estos casos es más sencillo aplicar la intervención sobre comunidades enteras (p. ej., que a una determinada localidad se añada flúor al suministro de agua y a otra no).*

- **Ensayo clínico no aleatorizado** (MIR): igual que un ensayo clínico, sólo que la asignación de la intervención no es aleatoria. También se denomina **estudio antes-después**, aunque este término se suele reservar para estudios de intervención en un único grupo de pacientes (con lo que no hay posibilidad de aleatorización) en los que se compara la situación basal con la situación tras la administración de un tratamiento (datos apareados).

(Ver figura 2)

7.3. Niveles de evidencia científica

Actualmente, las sociedades científicas médicas basan sus recomendaciones en los resultados de los estudios epidemiológicos disponibles sobre cada materia: **Medicina Basada en la Evidencia**, que consiste en la integración de la mejor evidencia científica disponible sobre un tema con la maestría clínica individual (MIR 10, 131).

Dichas recomendaciones se suelen recoger en **guías de práctica clínica**, que se elaboran por comités de expertos para las patologías más prevalentes que trate cada especialidad médica. En dichas guías, a cada recomendación que realizan los expertos se asocia, en general, una **clase de recomendación** y un nivel de evidencia científica. El **nivel de evidencia científica** indica la calidad de los estudios sobre los que se basa cada recomendación, mientras que la clase de recomendación se refiere al nivel de consenso que existe entre los redactores de las guías para indicar o contraindicar una determinada actitud (diagnóstica o terapéutica), en función de su relación beneficio/riesgo

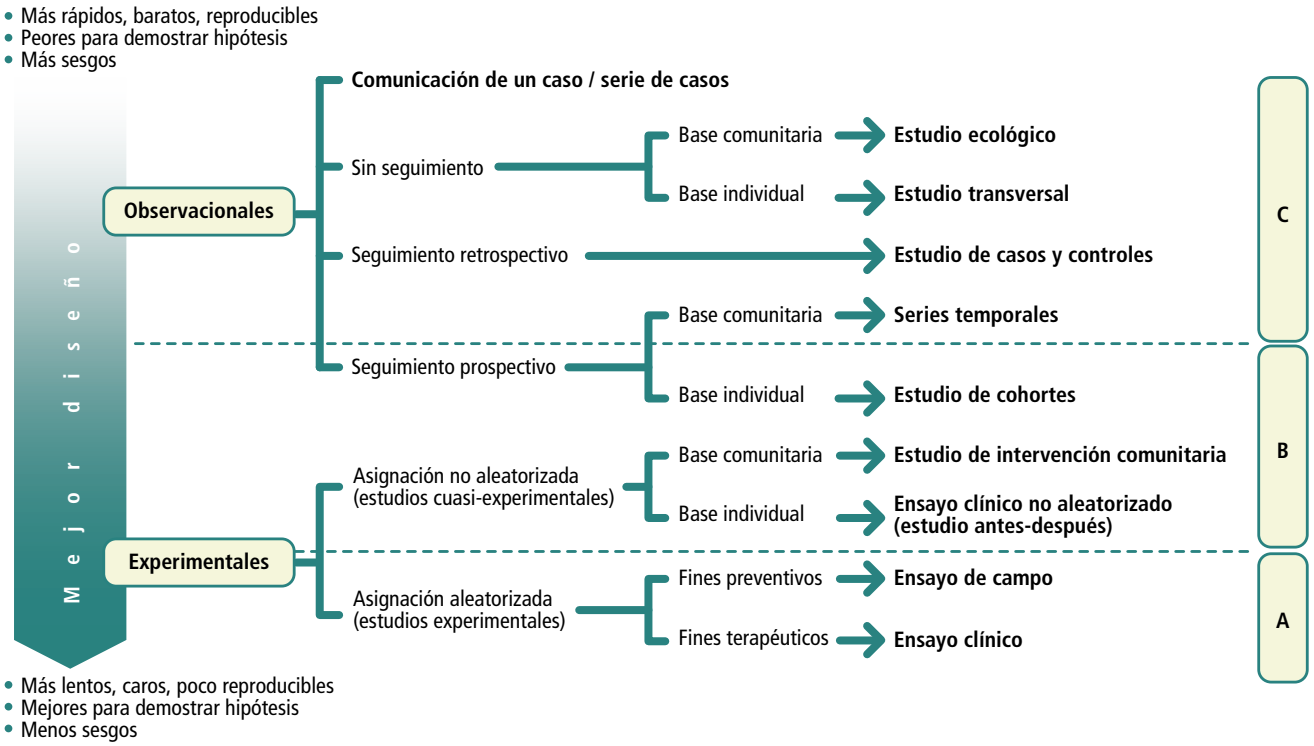


Figura 2. Tipos de estudios epidemiológicos y nivel de evidencia científica que generan.

según la evidencia científica disponible. Aunque es variable en función de cada sociedad científica, la mayoría de ellas utiliza la siguiente gradación para las clases de recomendación:

- **Clase I** (MIR 21, 182): hay consenso general y/o evidencia científica sobre el beneficio de la actitud que sea. "Está recomendada / está indicada".
- **Clase II**: hay divergencia de opiniones y/o evidencia no concluyente sobre el beneficio de la actitud que sea. Se divide en dos clases:
  - **Clase IIa**: existe una mayoría de estudios y de opiniones a favor del beneficio de la actitud. "Debería considerarse".
  - **Clase IIb**: la utilidad o beneficio está menos establecida según la evidencia y opinión de los redactores. "Podría considerarse".
- **Clase III**: hay consenso general y/o evidencia científica sobre la ausencia de beneficio, o incluso del perjuicio, de la actitud que sea. "No está recomendada".

Para determinar la clase de recomendación en una guía de práctica clínica se utilizan sistemas de gradación de la calidad de la evidencia y la fuerza de la recomendación como el sistema **GRADE** *(Grading of Recommendations, Assessment, Development and Evaluation)* (MIR 23, 46). Este sistema establece sus recomendaciones en función de 4 criterios: el balance entre beneficios y riesgos de la intervención (y la magnitud de la diferencia entre ambos); la calidad de la evidencia científica disponible; las preferencias de los pacientes; y el consumo de costes y recursos. Las recomendaciones pueden ser **a favor** o **en contra** de la intervención, y **fuertes** o **débiles** (en función de los 4 criterios mencionados).

Respecto al **nivel de evidencia científica**, en función de la calidad de diseño de cada tipo de estudio, éste es capaz de generar un determinado nivel de evidencia científica. Existen numerosas escalas de cuantificación del nivel de evidencia, pero la más utilizada utiliza un esquema **ABC** (siendo mayor la evidencia generada por los estudios de nivel A, y menor la de nivel C, cuya evidencia se considera inconcluyente y debe confirmarse mediante estudios de mayor calidad).

Niveles de evidencia científica ABC (MIR 11, 176; MIR)

- **Nivel de evidencia A**.
  - Metaanálisis de estudios experimentales aleatorizados.
  - Varios estudios experimentales aleatorizados.
- **Nivel de evidencia B**.
  - Un único estudio experimental aleatorizado.
  - Estudios cuasi-experimentales.
  - Estudios de cohortes grandes.
- **Nivel de evidencia C**.
  - Estudios observacionales (salvo cohortes).
  - Consenso de expertos.

Metaanálisis

El metaanálisis es una **revisión sistemática** de la literatura en la que **se combinan estadísticamente** los resultados de todos los estudios incluidos (MIR 14, 194; MIR). La colaboración Cochrane es una organización sin ánimo de lucro que promueve las revisiones sistemáticas, y que ha realizado guías sobre cómo realizarlas con calidad.

- **Revisión sistemática**: la búsqueda bibliográfica se realiza en función de unos **criterios de selección** concretos (inclusión y exclusión), de modo que todo estudio que cumpla con esos criterios deberá ser incluido en el metaanálisis (MIR 18, 219). *Las **revisiones narrativas**, sin embargo, incluyen los artículos que elija libremente el investigador*.

La colaboración Cochrane dispone de una herramienta, en los metaanálisis de ensayos clínicos, para valorar la calidad de cada estudio individual que se está evaluando para su inclusión, y que puntúa los siguientes ítems:

- Aleatorización de la intervención (evitar sesgo de selección). Para una correcta aleatorización, la secuencia de asignación de la intervención debe generarse aleatoriamente (obviamente), y dicha secuencia no debe ser conocida por los investigadores.
- Enmascaramiento de pacientes e investigadores (evitar sesgo de clasificación de tipo "rendimiento").
- Enmascaramiento de la evaluación de los eventos que aparezcan durante el seguimiento (evitar sesgo de clasificación de tipo "detección").
- Que los datos de eventos clínicos que ocurran en el seguimiento no estén incompletos debido a pérdidas en el seguimiento (evitar sesgos de atricción).
- Que todos los resultados importantes se hayan reportado, y no solo aquéllos que interesen a los investigadores (evitar sesgo de información científica).

Para explicar gráficamente los pasos de la búsqueda bibliográfica, se suele utilizar el diagrama de flujo PRISMA **(ver figura 3)**.

- **Combinación estadística de resultados**: los resultados de los pacientes que participaron en cada estudio individual se tratan como si todos los pacientes hubieran participado en un único estudio. Así, se obtiene un **resultado combinado** que procede de un tamaño muestral inmenso (la suma de tamaños muestrales de cada estudio individual), con lo que se consigue:
  - **Mayor precisión**: intervalos de confianza más pequeños.
  - **Mayor potencia estadística**: mayor probabilidad de demostrar la existencia de diferencias de modo estadísticamente significativo, si realmente existen.

Los metaanálisis se realizan en situaciones en las que la **evidencia científica** disponible sobre un tema es **inconcluyente**, bien porque los estudios realizados tienen un pequeño tamaño muestral y no han demostrado diferencias significativas (quizá por falta de potencia estadística), o bien porque los resultados de los distintos estudios arrojen conclusiones discordantes.

La evidencia generada por los metaanálisis se considera **superior** a la de los estudios individuales incluidos **(MIR 17, 234)**, y ocupa el nivel ABC del estudio incluido que tenga un menor nivel de evidencia. *Por ejemplo, un metaanálisis de ensayos clínicos tendrá un nivel de evidencia A, pero si se mezclan ensayos clínicos con estudios de cohortes el nivel de evidencia será B.*

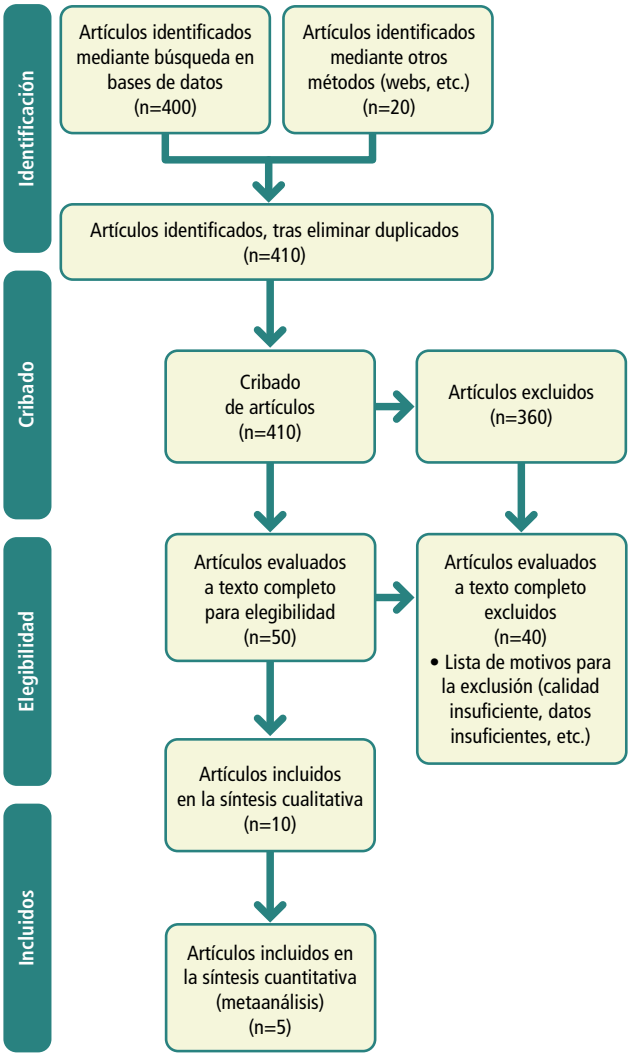


Figura 3. Diagrama de flujo PRISMA para las revisiones sistemáticas. Inicialmente se realiza una búsqueda bibliográfica (en bases de datos fundamentalmente, pero también en otras fuentes) y se seleccionan artículos que por su título podrían cuadrar con los criterios de inclusión y exclusión de nuestro estudio. Se realiza entonces un cribado (*screening*) de dichos estudios leyendo su abstract; en dicho cribado la mayoría de artículos quedarán eliminados, y quedarán unos pocos para realizar una revisión más minuciosa leyendo el texto completo. De dichos artículos, solo algunos serán finalmente incluidos en la revisión sistemática, y solo unos pocos de ellos podrán utilizarse para la agrupación estadística de resultados (metaanálisis).

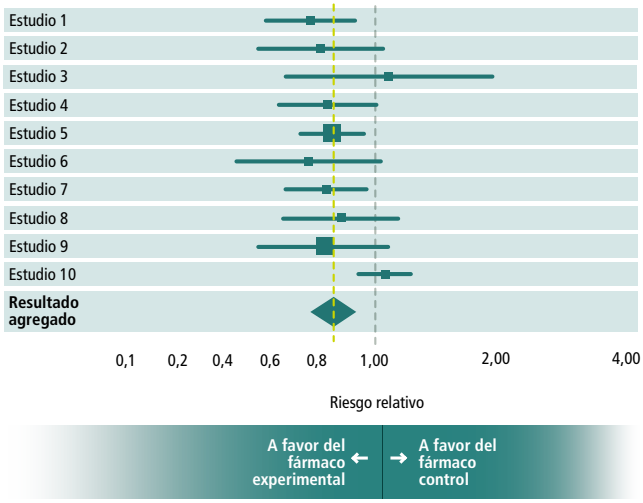


Figura 4. Diagrama de bosque o *forest plot* **(MIR 21, 46)**. El *forest plot* es un método de representación gráfica de los resultados de un metaanálisis. A su izquierda aparecen listados cada uno de los estudios que se han incluido en el metaanálisis, y a su derecha aparecen representados los resultados mediante un cuadrado (estimador de riesgo del estudio, por ejemplo el RR o el OR **(MIR 23, 48)**) y una línea horizontal (IC 95% del resultado del estudio). Debajo de los estudios individuales se expone el IC 95% del resultado agregado tras realizar el metaanálisis, representado habitualmente como un rombo cuyo centro coincide con el del IC 95% del resultado agregado. El tamaño de cada cuadrado (estimador de riesgo) vendrá determinado por el peso respectivo de cada estudio en el metaanálisis, que a su vez depende habitualmente del tamaño muestral del estudio. Se marca con una línea vertical el límite de no significación (si los IC 95% de cada estudio o el resultado agregado cruza esa línea, es que los resultados no son significativos).

Aspectos estadísticos del metaanálisis (MIR 13, 194)

Heterogeneidad y modelos estadísticos de combinación de resultados

Uno de los principales problemas que nos podemos encontrar en un metaanálisis es la diferencia entre los estudios incluidos. Para determinar el grado de heterogeneidad de los estudios incluidos se utilizan diversos métodos, como los estadísticos de heterogeneidad (prueba Q, índice I2) y el gráfico de Galbraith. Cuando encontremos un análisis de heterogeneidad con una p significativa (<0,05), significará que existe heterogeneidad, y por tanto lo correcto será aplicar un modelo de efectos aleatorios (MIR 16, 28).

Para combinar los estudios y obtener el resultado agrupado global del metaanálisis, se calcula la media del resultado de los estudios individuales, ponderados según el tamaño muestral, la dispersión y la calidad de los mismos (MIR 22, 43), dando diversos grados de importancia a cada factor en los diferentes modelos. El factor principal que nos hace elegir el modelo estadístico de combinación de los resultados es el grado de **heterogeneidad** de los estudios, existiendo un modelo de efectos fijos y uno de efectos aleatorios.

El **modelo de efectos fijos** se utiliza cuando la heterogeneidad es baja (MIR 11, 195); no tiene en cuenta la variabilidad entre los estudios (inter-estudios), sino que su ponderación se basa principalmente en la variabilidad intra-estudio y en el tamaño muestral. El modelo de efectos fijos se utiliza por tanto cuando los estudios incluidos son muy parecidos entre sí y con el mismo tipo de pacientes (similar a un ensayo clínico con criterios de selección muy estrictos), lo que le confiere mayor validez interna y una mayor potencia estadística y precisión en sus resultados.

Por otro lado, el **modelo de efectos aleatorios** tiene en cuenta que los estudios pueden ser variables entre sí (variabilidad inter-estudios). De este modo, los modelos de efectos aleatorios son menos potentes, con intervalos de confianza más amplios para el efecto combinado, y al no ponderarse por tamaño muestral, pueden dar excesiva importancia a estudios de pequeño tamaño.

Análisis de sensibilidad

Este análisis pretende estudiar la influencia de cada uno de los estudios incluidos en la estimación global del efecto y así la estabilidad de la medida final. Consiste en la repetición del metaanálisis tantas veces como estudios se hayan incluido, de forma que cada vez se omite un estudio combinándose todos los restantes. Si los resultados de los distintos metaanálisis realizados son similares, se puede concluir que los resultados son robustos. En caso contrario no se tendría un estimador robusto, lo cual exigiría cierta precaución en la interpretación de los resultados o podría ser motivo para generar nuevas hipótesis.

Este mismo proceso podría repetirse eliminando a un mismo tiempo varios estudios (por ejemplo, aquellos de peor calidad metodológica, los no publicados, etc.) para determinar su posible influencia en los resultados.

Sesgo de publicación

Es un tipo de sesgo de selección que consiste en no incluir artículos que no se hayan publicado (si no los buscamos en sitios muy concretos nunca vamos a encontrar esos artículos). Dichos artículos suelen ser desfavorables para el tratamiento experimental (simplemente porque los estudios en los que el fármaco no muestra diferencias significativas no se suelen publicar); por tanto, si existe sesgo de publicación **sobreestimaremos el efecto del fármaco experimental**.

Existen varios métodos para investigar la presencia de un sesgo de publicación. El más simple consiste en realizar un **análisis de sensibilidad** para calcular el número de estudios negativos realizados y no publicados que debería haber para modificar el sentido de una posible conclusión "positiva" del metaanálisis (si este número es muy elevado, se considera que la probabilidad de que el sesgo de publicación haya modificado los resultados es baja, y se acepta la existencia de las diferencias sugeridas por el metaanálisis).

También se puede examinar con el método conocido como el gráfico en embudo (Funnel-plot) (MIR 15, 200) (ver figura 5), en el que se distribuyen los estudios incluidos en el metaanálisis. Se parte del supuesto de que los estudios con mayor probabilidad de no ser publicados son los que no muestran diferencias (estudios "negativos"), sobre todo si son de pequeño tamaño; si no encontramos en nuestro gráfico estudios con esas características, es probable que estemos incurriendo en un sesgo de publicación. Como este gráfico presenta una subjetividad "visual", en ocasiones se utilizan también pruebas analíticas para verificar el sesgo de publicación, como el test de Begg o el test de Egger.

Metaanálisis acumulado

Se define como aquel proceso en el cual se lleva a cabo un nuevo metaanálisis cada vez que aparece un nuevo estudio publicado. No requiere de técnicas estadísticas especiales para combinar los estudios. Permite estudiar de forma retrospectiva el momento en el que el efecto de un nuevo tratamiento supera al control.

Esta forma de presentación de los resultados pone de manifiesto lo difícil que es para cualquier estudio individual, una vez se han alcanzado resultados relativamente estables, aportar información adicional.

El metaanálisis acumulado sería similar en planteamiento a un ensayo clínico secuencial.

7.4. Estructura metodológica de un trabajo científico (MIR 18, 225)

Un proyecto de investigación debe situar las bases de la investigación a realizar. Su valor se establece en la medida en que tiene plena claridad y concreción en las razones para analizar el objeto de estudio elegido, la perspectiva teórica desde donde se sitúa el investigador, el paradigma investigativo que sustenta todo el estudio y, por tanto, la metodología de aproximación a la realidad: población, muestra, estrategias de recogida de información, técnicas de análisis de la información y temporalidad de todo el proceso. Por orden, los elementos fundamentales en esta estructura son:

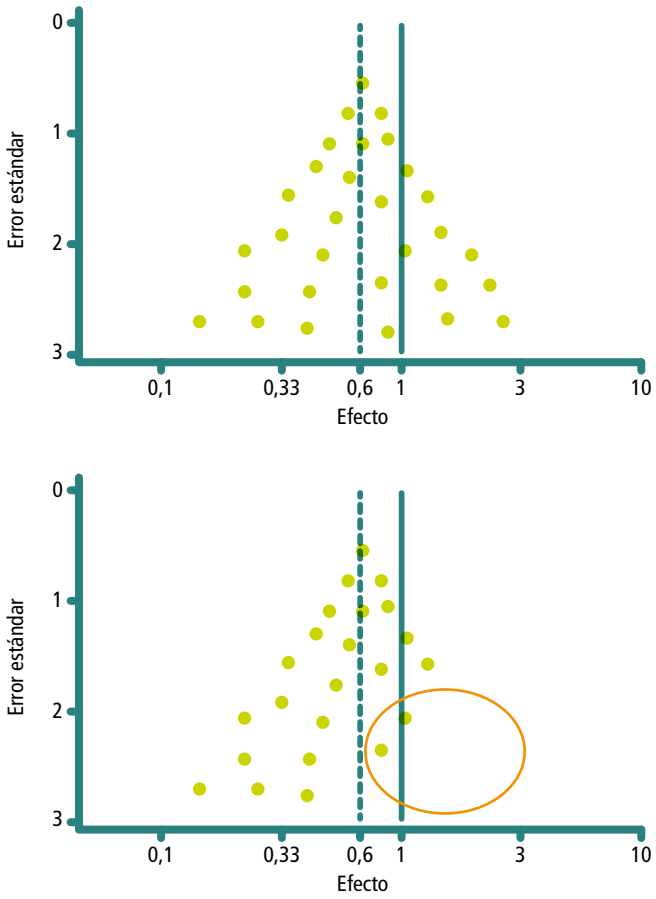


Figura 5. *Funnel plot* **(MIR 17, 18)**. El gráfico representa cada estudio (puntos) incluido en el metanálisis en función de su tamaño (que es inversamente proporcional al error estándar del resultado, mostrado en el eje de ordenadas: estudios pequeños "abajo" y estudios grandes "arriba") y de la medida del efecto que se obtuvo (RR, OR, etc., representado en el eje de abscisas). La medida del efecto agregado del metanálisis suele representarse como una línea vertical punteada. Si la distribución de los estudios sigue una forma triangular y simétrica (esquema de arriba) a ambos lados de dicha medida de efecto agregado, existen pocas probabilidades de sesgo de publicación, pero si "faltan" estudios en alguna región del teórico triángulo (esquema de abajo, círculo rojo) es probable la presencia de un sesgo de publicación.

1. **Introducción**: contiene una descripción clara de la estructura general del proyecto.
2. **Justificación**: contiene los argumentos fundamentales que sustentan la investigación a realizar.
3. **Planteamiento del problema**: formulación del problema que se pretende resolver con la investigación.
4. **Objeto de estudio**: delimita la parte de la realidad que interesa estudiar.
5. **Preguntas de investigación**: son las interrogantes básicas que se derivan de la justificación y el problema planteado.
6. **Objetivos**: las acciones concretas que se realizarán para intentar responder a las preguntas de investigación. En este apartado se incluye la **hipótesis** de investigación.
7. **Fundamentación teórica**: directrices teóricas que guían el estudio, con las evidencias de la literatura.
8. **Metodología de la investigación**: descripción y argumentación de las principales decisiones metodológicas.
9. **Población y muestra**: selección de la población objetivo. Justificación del tamaño muestral elegido.
10. **Diseño de la investigación**: representación gráfica que presenta la metodología completa, la forma en que se organiza todo el proceso de investigación y los aspectos metodológicos esenciales.
11. **Cronograma o carta Gantt y presupuesto**: estimación del tiempo y dinero que tomarán cada una de las etapas de la investigación.
12. **Bibliografía**: fuentes documentales consideradas, cumpliendo las normas estandarizadas (p. ej., el estilo Vancouver).

7.5. Fases de realización de los estudios epidemiológicos

1. Diseño

Idealmente, todo proyecto de investigación debería iniciarse con una pregunta. Se ha propuesto el acrónimo **PICO**, que aúna los componentes que debe tener toda pregunta que guía una investigación **(MIR 20, 150)**:

- **P**acientes: criterios de inclusión/exclusión. Cuando realizamos una revisión sistemática, los criterios de inclusión y exclusión son referidos a los estudios que seleccionaremos (que pasan a ser nuestros "pacientes").
- **I**ntervención: tratamiento en el brazo experimental.
- **C**omparación: tratamiento en el brazo sin intervención.
- ***O**utcome* (resultado): resultado en el que nos fijaremos para evaluar nuestra intervención; idealmente debe contener uno o dos objetivos primarios, y otros secundarios.

Además, antes de empezar el estudio (***a priori***), se deben especificar todos sus aspectos metodológicos, con un cuidado especial en explicar todas las mediciones que se van a realizar y el método estadístico por el que se van a analizar los resultados. Cuando se realiza un ensayo clínico, existe obligación de incluirlo en registros públicos (p. ej. clinicaltrials.gov) una vez se completa la fase de diseño, para poder evaluar posteriormente que efectivamente se hayan seguido todos los aspectos metodológicos incluidos en el diseño.

Si se extraen conclusiones de un estudio epidemiológico **a posteriori (estudios *post hoc*)**, dichas conclusiones no servirán para confirmar hipótesis, sino que sólo sirven para generarlas y se deberán confirmar con nuevos estudios específicamente diseñados para ello. *Por ejemplo: realizar un análisis de subgrupos que no estaba planificado inicialmente y observar que un determinado subgrupo se beneficia del fármaco.*

2. Reclutamiento (inclusión de sujetos participantes)

Una vez definidas las características del estudio, se realiza la inclusión de los sujetos participantes. Para incluir a un paciente en un estudio experimental, deben seguirse **por este orden** (MIR) los siguientes pasos:

A. **Criterios de selección**: verificar que el paciente cumpla todos los **criterios de inclusión** y no tenga ningún **criterio de exclusión**.

B. **Consentimiento informado** (MIR 18, 208): el paciente debe expresar libremente y por escrito su consentimiento para participar en el estudio.

   En el caso de que se utilice **placebo**, los sujetos deben saber que pueden ser tratados con éste (MIR), aunque luego no podrán conocer si están tomando el placebo o el fármaco (enmascaramiento). Además, si se utiliza placebo deben incluirse en el diseño del estudio "cláusulas de rescate" que permitan pasar a un sujeto del grupo placebo al grupo de fármaco experimental si su evolución clínica empeora claramente y hay datos provisionales de mayor eficacia con el fármaco experimental.

   Cuando administramos un fármaco en un ensayo clínico, su efecto farmacológico total (respuesta global) tendrá los siguientes componentes (MIR 19, 117):

   - Efecto farmacodinámico: es el efecto real terapéutico del fármaco.
   - Efecto placebo: es la suma de varios efectos distintos:
     - Efecto placebo absoluto: se debe a la suma del efecto inespecífico del fármaco (el efecto placebo que supone recibir un fármaco) y al efecto inespecífico del médico (el efecto placebo que supone sentirse tratado por un médico).
     - Regresión a la media: si cuando administramos el fármaco a un sujeto concreto de un estudio su efecto se aleja mucho de la media (efecto extremo: o muy alto o muy bajo), en el siguiente sujeto, o la siguiente vez que se administre el fármaco al mismo paciente, el efecto tenderá a ser más cercano a la media (dado que es el efecto más probable de encontrar).

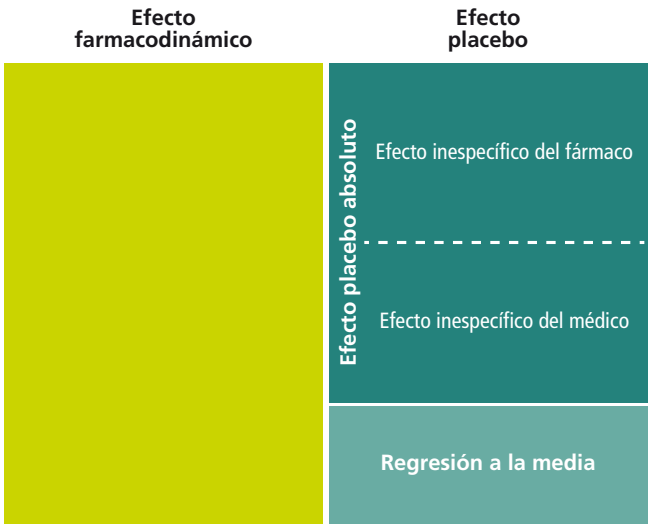


Figura 6. Componentes del efecto farmacológico total.

C. **En estudios experimentales: asignación de intervención**. Una vez se comprueba que el paciente cumple los criterios de selección y firma el consentimiento informado, se le asigna a un grupo del estudio.

   En los estudios experimentales "puros" esta fase se denomina **aleatorización** (MIR 10, 189). Si la aleatorización ha sido adecuada, las características de los distintos grupos del estudio serán superponibles, pero el azar también puede jugar una mala pasada y hacer que la distribución de alguna característica sea distinta entre los grupos (MIR). Existen distintas técnicas de aleatorización:

   - Aleatorización simple: cada paciente tiene las mismas probabilidades de formar parte de cada uno de los grupos. El azar podría hacer que el número de pacientes de cada grupo sea distinto.
   - Aleatorización por bloques: consigue que haya el mismo número de personas en todos los grupos de estudio. *Para ello, se elige un número, y cada "bloque" (ese número) de pacientes debe tener el mismo número de pacientes distribuido en cada grupo de tratamiento.*
   - Aleatorización estratificada: la estratificación de la muestra en función de una característica que pueda funcionar como factor de confusión o que sea un factor pronóstico importante permitirá su distribución homogénea en los distintos grupos (MIR).

3. Monitorización

Fase de **seguimiento** de los pacientes, en la que se realizan las mediciones previstas en el diseño y se obtienen los resultados.

4. Análisis de resultados y obtención de conclusiones (MIR 14, 204; MIR)

Una vez obtenidos los resultados, se deben analizar siguiendo los siguientes pasos:

A. Verificar que el **diseño** sea correcto y no haya sesgos (**validez interna**). Es el paso **más importante** del análisis de resultados, ya que la ausencia de validez interna invalidará cualquier resultado obtenido.

   La lectura crítica de un artículo científico deberá, por tanto, basarse fundamentalmente en analizar los aspectos metodológicos del estudio (su diseño), plasmados en el apartado "Material y Métodos" del estudio.

B. Confirmar la existencia de significación estadística.

C. Valorar la magnitud de las diferencias existentes y su relevancia clínica. La existencia de diferencias estadísticamente significativas entre dos intervenciones no implica que una de ellas sea mejor que otra. Sólo si la magnitud de dicha diferencia es suficiente (MIR), y si implicará beneficios relevantes desde un punto de vista clínico, podremos decir que una intervención es "mejor" que otra y, por tanto, establecer una recomendación al respecto (MIR).

D. Determinar la validez externa. Los resultados del estudio se podrán generalizar a aquellos subgrupos poblacionales que cumplan los criterios de selección de los pacientes participantes del estudio.

5. Difusión de los resultados

Mediante la publicación del estudio en artículos científicos u otros medios de comunicación.

7.6. Fases de desarrollo de un tratamiento (fases del ensayo clínico) (MIR)

Para el desarrollo de un tratamiento, primero se realiza su síntesis química y se realizan estudios con material biológico in vitro. Posteriormente se prueba su eficacia y seguridad en animales. Tras esta **fase preclínica**, el tratamiento debe testarse en humanos antes de poder comercializarlo.

La **fase clínica** del desarrollo de un tratamiento incluye estudios con diseño de ensayo clínico y por eso se suele hablar directamente de "fases del ensayo clínico"; dicha fase clínica consta a su vez de las siguientes fases:

Fase I (MIR 15, 179; MIR)

Es la primera vez que se utiliza el tratamiento en humanos (MIR). Consiste en un estudio transversal sobre un único grupo de sujetos **voluntarios**, cuyo objetivo es estudiar las **propiedades farmacocinéticas** del tratamiento.

Como objetivo secundario, se estudia de forma preliminar la **tolerabilidad/toxicidad** del fármaco (si su administración produjo efectos adversos a los voluntarios) (MIR 11, 188).

Habitualmente se realiza sobre **voluntarios sanos** (que suelen recibir una remuneración económica por participar en el estudio), pero si el tratamiento se prevé que tenga muchos efectos adversos se realizará con voluntarios enfermos (MIR) *(p. ej., quimioterápicos)*.

Fase II y fase III

Son las fases que tienen **diseño de ensayo clínico**. Por lo tanto, su objetivo es valorar la **eficacia y seguridad** del tratamiento en un grupo de **enfermos**. Las dos fases pueden tener un grupo control y un diseño idéntico. En el caso de tener grupo control, éticamente se debe emplear el **fármaco activo** que sea actualmente de elección (MIR 12, 185; MIR 12, 189). La utilización de **placebo** se reserva para los casos en los que no hay fármacos con eficacia demostrada, y también se permite para enfermedades que no sean graves (sin riesgo de secuelas) y en las que exista una alta tasa de respuesta a placebo (MIR 14, 202; MIR 12, 232).

Las diferencias entre la fase II y la fase III son las siguientes:

Fase II (MIR 21, 48; MIR 14, 203; MIR 13, 190)

Empleo del tratamiento en un **grupo reducido** de enfermos (en general <100) muy seleccionados, esto es, con **criterios de selección estrictos (estudios explicativos)** (MIR). En esta fase se evalúan adecuadamente datos de eficacia preliminares, pero la evaluación de la seguridad es más difícil, ya que al tener un pequeño tamaño muestral existen menos probabilidades de observar efectos adversos (que suelen ser infrecuentes).

Además, suelen utilizar **variables resultado "blandas"** (MIR) (determinaciones de laboratorio, pruebas de imagen...), que aportan menor relevancia clínica de los resultados, pero al ser en general cuantitativas proporcionan mayor potencia estadística.

En ocasiones se distingue una primera fase IIa (la que hemos mencionado hasta ahora), y una segunda **fase IIb** (MIR 17, 116; MIR 10, 187), cuyo objetivo es probar varias dosis del fármaco para establecer la relación dosis-respuesta (**titulación de dosis**) y elegir la dosis más adecuada para su empleo en la fase III.

La utilización de **criterios de selección estrictos** hace que la muestra sea muy homogénea (los pacientes se parecerán mucho entre sí), lo cual confiere las siguientes ventajas e inconvenientes:

- **Ventajas**:
  - Mayor **validez interna**: menor riesgo de sesgos de selección o por factor de confusión, al tener los distintos grupos características similares.
  - **Resultados más homogéneos**: como los pacientes son muy parecidos, el fármaco hará lo mismo en todos ellos obteniendo resultados más **precisos** (intervalos de confianza más pequeños), con lo que aumenta la **potencia estadística** del estudio (necesidad de menor tamaño muestral).
- **Inconvenientes**:
  - Menor **validez externa**: los resultados serán generalizables sólo a un limitado sector de la población.

Fase III (MIR 13, 191; MIR)

Empleo del tratamiento en un **grupo amplio** de enfermos (en general >100) con **criterios de selección laxos (estudios pragmáticos)**, lo que va a permitir que la muestra sea similar a los pacientes que se van a encontrar en la práctica clínica habitual (MIR 15, 180), y la eficacia demostrada se parecerá a la "efectividad" que se observará en la población (MIR).

Además, suelen utilizar **variables resultado "duras"** (variables clínicas: muerte, infarto, ictus...), que aportan **mayor relevancia clínica** de los resultados, pero al ser en general cualitativas *(p. ej., muerte: sí/no)* proporcionan menor potencia estadística. Estas variables clínicas son muchas veces **subjetivas** (dolor, calidad de vida...), lo cual dificulta su correcta determinación; son útiles en este caso las **escalas de evaluación multidimensionales**, que aportan una alta relevancia clínica, pero pueden ser difíciles de interpretar y tener problemas de validez para nuestra muestra concreta (MIR).

En ocasiones, en vez de utilizar variables duras se utilizan variables resultado blandas pero que han demostrado asociarse de manera significativa a una determinada variable dura en estudios previos ("validadas"): es lo que se llaman **variables subrogadas o intermedias** (MIR). Al ser variables en general cuantitativas, aportarán mayor potencia estadística y por tanto permitirán utilizar un menor tamaño muestral en el estudio, pero son siempre peores que la utilización de la variable dura en sí misma (más riesgo de sesgos). *Ejemplo: analizar si un nuevo betabloqueante consigue disminuir la PAD en 10 mmHg (variable subrogada) en sujetos hipertensos, lo cual ha demostrado en estudios previos aumentar la supervivencia (variable dura).*

Así, los estudios de fase III demuestran si el nuevo tratamiento va a ser útil para los pacientes desde el punto de vista clínico y constituyen la evidencia fundamental del beneficio-riesgo del medicamento, por lo que son los que permiten que se comercialice un fármaco o tratamiento.

La utilización de criterios de selección laxos hace que la muestra sea muy heterogénea (los pacientes serán muy diferentes entre sí), lo cual confiere las siguientes ventajas e inconvenientes:

- **Ventajas**:
  - Mayor **validez externa**: los resultados serán generalizables a un amplio sector de la población *(así, cuando se comercialice el fármaco una gran parte de la población se podrá beneficiar de él)*.
- **Inconvenientes**:
  - Menor **validez interna**: mayor riesgo de sesgos de selección o por factor de confusión.
  - **Resultados dishomogéneos**: como los pacientes son distintos entre sí, el efecto del fármaco variará mucho en función de sus características, obteniendo resultados **menos precisos** (intervalos de confianza amplios) y una **menor potencia estadística** (necesidad de un mayor tamaño muestral).

Fase IV (fase poscomercialización)

Incluye aquellos estudios realizados con un fármaco tras su comercialización, que como hemos visto ocurre tras la publicación de los estudios en fase III.

Tiene fundamentalmente tres **objetivos**:

- Estudio de la **efectividad** de un fármaco (cuando se utiliza en la práctica clínica habitual) **(MIR 10, 192)**. Para dicho objetivo, se realizan EPA: estudios postautorización de tipo observacional y seguimiento prospectivo.
- Búsqueda de **nuevas indicaciones**: para ello, se deberán volver a realizar estudios con diseño análogo a la fase II y fase III.
- **Farmacovigilancia** **(MIR 17, 119; MIR)**: sistema de notificación espontánea de posibles reacciones adversas asociadas a un tratamiento por parte del personal sanitario. Intenta detectar **reacciones adversas poco frecuentes** (que no se detectaron en los ensayos clínicos realizados por su limitado tamaño muestral) y que sólo se podrán detectar cuando el fármaco se administre a miles de personas en la práctica clínica habitual.

  La notificación espontánea de reacciones adversas la debe realizar cualquier personal sanitario mediante la cumplimentación de una "**tarjeta amarilla**" **(MIR)** que se envía a la Agencia Española del Medicamento y Productos Sanitarios (AEMPS), o bien a través de la aplicación web y móvil FEDRA. Ésta recopila las notificaciones recibidas y las reenvía a la Agencia Europea del Medicamento, que es el organismo responsable de la Farmacovigilancia a nivel europeo.

  Si se reciben varias notificaciones que parecen sugerir que un medicamento produce una reacción adversa, se lanza una advertencia al respecto ("warning") y se diseña un estudio epidemiológico para demostrar la posible relación causal. Al querer estudiar un problema de salud raro (reacción adversa poco frecuente) el diseño habitual de dichos estudios es de **casos y controles**.

Los **estudios farmacoeconómicos** **(ver tema 7.10. Estudios farmacoeconómicos)** son útiles para labores de farmacovigilancia porque permiten estudiar la utilización de un fármaco en la población y por tanto calcular cuántas personas están expuestas al mismo **(MIR)**.

FASE	POBLACIÓN	CARACTERÍSTICAS	OBJETIVOS
I	Voluntarios (sanos)	Diseño transversal	• Características farmacocinéticas • Toxicidad preliminar
II	Enfermos	• Criterios de selección estrictos • ↓ n • **Variables blandas**	• **IIa**: eficacia y seguridad • **IIb**: titular dosis
III	Enfermos	• Criterios de selección laxos • ↑ n • **Variables resultado duras**	Eficacia y seguridad
IV	Práctica clínica habitual		• Efectividad • Nuevas indicaciones • Farmacovigilancia

Tabla 2. Fases del ensayo clínico.

7.7. Diseños especiales en estudios experimentales

Diseño paralelo *vs.* diseño cruzado

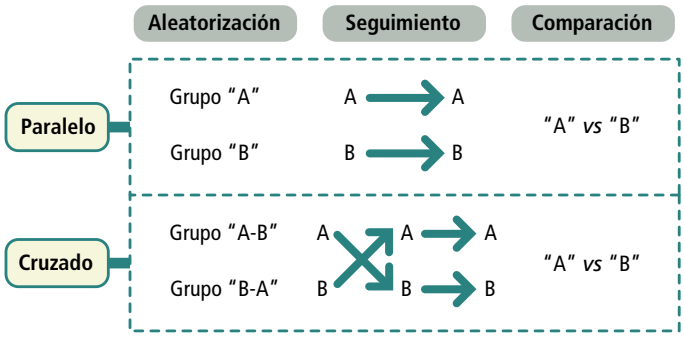


Figura 7. Diseño paralelo y diseño cruzado.

El **diseño paralelo** es el utilizado habitualmente, por el cual un grupo de sujetos recibe un único tratamiento o intervención, y el otro grupo que recibe el otro tratamiento.

En el **diseño cruzado**, sin embargo, todos los sujetos reciben los dos tratamientos o intervenciones en comparación **(MIR)**. Un grupo recibe primero un tratamiento y luego el otro, y el otro grupo recibe la secuencia de tratamientos inversa. Así, lo que se aleatoriza en este caso es la secuencia de administración de los tratamientos **(MIR)**.

Ventajas del diseño cruzado respecto al paralelo

- **Requiere la mitad de tamaño muestral**. En un diseño cruzado, cada individuo recibe los dos tratamientos, por lo que sirve como su propio control y ahorra la inclusión de otro individuo para servir de control. *Si necesitamos 100 pacientes que tomen la medicación "A" y 100 pacientes que tomen la medicación "B", en un diseño paralelo tendremos que incluir 200 pacientes (100 por ramo de tratamiento), mientras que en un diseño cruzado sólo necesitaremos 100 pacientes (50 en el grupo A-B y 50 en el grupo B-A).*
- **Muestra más homogénea**. El grupo de individuos que toma el fámaco "A" es igual que el que toma el fármaco "B" (son las mismas personas); esta circunstancia hace surgir el concepto de "**variabilidad intraindividual o intragrupal**" **(MIR)**, que hace referencia al hecho de que en los diseños cruzados cada individuo es su propio control y por tanto la variabilidad que existe entre cada individuo y su control (o entre cada grupo y su control) es nula (son las mismas personas). Sin embargo, la "variabilidad interindividual o intergrupal" (entre dos sujetos o grupos diferentes del estudio) no varía respecto al estudio paralelo.

  Como ya hemos visto previamente, el disponer de una muestra homogénea aporta ventajas:

  - Menor riesgo de sesgos de selección y por factor de confusión.
  - Mayor potencia estadística y precisión de los resultados.

Inconvenientes del diseño cruzado respecto al paralelo

- **Mayor duración**: dura el doble de tiempo, ya que cada sujeto recibe una intervención y luego debe recibir la otra.
- **Muy sensible a las pérdidas**: si se pierde un paciente, se pierde en los dos grupos de comparación "A" y "B", por lo que es como perder dos pacientes en un diseño paralelo.
- **Efecto residual o de arrastre**: efecto que deja el primer fármaco sobre el organismo. Tras dejar de tomar el primer fármaco, se debe esperar un cierto tiempo (**periodo de lavado**) para que se elimine el fármaco y desaparezca su efecto residual.
- Por lo tanto, el tratamiento **no puede ser curativo ni dejar un efecto irreversible** **(MIR 17, 117; MIR)**. Además, el **periodo de lavado** de los dos fármacos debe ser **similar** **(MIR)**.
- **Efecto periodo**: los dos tratamientos se administran al paciente en dos periodos de tiempo distintos, por lo que las características clínicas de la enfermedad no deben cambiar entre esos dos periodos para poder comparar el efecto de los dos fármacos en las mismas condiciones basales.

  Por lo tanto, la enfermedad debe ser **crónica**. No es útil para enfermedades agudas o que cursen con brotes, a menos que los brotes sean predecibles **(MIR)** *(si los brotes son predecibles es el escenario más favorable para utilizar un diseño cruzado, ya que el periodo entre los brotes funciona como periodo de lavado)*.

Diseño secuencial

No existe un tamaño muestral predeterminado, sino que se van incluyendo progresivamente pacientes hasta alcanzar un tamaño muestral o un periodo de tiempo máximo establecido **(MIR)**.

A medida que se van incluyendo pacientes, se realizan análisis intermedios para ver si se consigue llegar a la significación estadística. En el momento que se alcance la significación (o si se llega al tamaño muestral o el tiempo máximos), el estudio se detiene.

Diseño factorial

Es el diseño más **eficiente** cuando existen **>2 opciones de tratamiento** **(MIR 15, 178; MIR)**. Consiste en dividir la muestra en grupos que toman cada tratamiento por separado, y grupos que toman cada una de las posibles combinaciones de tratamientos.

Ventajas

- Permite evaluar **interacciones** entre los tratamientos.
- **Ahorra tamaño muestral**: ya que los pacientes que toman varios tratamientos cuentan para los resultados de cada uno de los tratamientos.

  Ejemplo: en un estudio que compara los fármacos A, B, C, en el que se necesitan 100 pacientes por rama, un diseño convencional utilizaría 300 pacientes (100 A, 100 B, 100 C), mientras que el siguiente ejemplo de diseño factorial utilizaría 220 pacientes (50 A, 50 B, 50 C, 20 AB, 20 AC, 20 BC, 10 ABC; sigue habiendo 100 pacientes en total que toman A, 100 que toman B y 100 que toman C).

Diseño con n = 1 (MIR 10, 193)

Ensayo clínico de **diseño cruzado** (sus características son totalmente aplicables) con un único individuo como muestra. Se realiza cuando un paciente crónico es refractario a los tratamientos habituales, y su objetivo es encontrar un tratamiento que le sea útil desde el punto de vista clínico (se evalúan **variables duras**).

A diferencia del resto de estudios epidemiológicos, el objetivo del estudio es mejorar la salud de nuestro paciente, en lugar de mejorar la salud de la población.

Diseño polietápico

Los tratamientos se administran primero con dosis de inducción, y posteriormente con dosis de mantenimiento (en varias "etapas"). Se utiliza principalmente en tratamientos antineoplásicos.

Utilización de controles históricos

Consiste en utilizar como grupo control a pacientes que han sido tratados de la patología que estamos investigando **en el pasado** (utilizándose por lo tanto el tratamiento convencional que hubiera disponible). Compararemos los resultados de estos pacientes con los resultados de un grupo de enfermos del presente a los que tratamos con la terapia experimental. La utilización de controles históricos tiene numerosas **limitaciones**:

- Utilización de **datos indirectos** (utilizamos datos de pacientes que fueron tratados en el pasado por otros médicos; pueden faltar datos que necesitemos, o pueden estar recogidos de un modo distinto al que nos interesa).
- Posibilidad de **sesgos de cointervención** (MIR 18, 228; MIR). Las "cointervenciones" son todas las mejoras en el manejo de la enfermedad que han aparecido entre el pasado y la actualidad, aparte del tratamiento experimental (p. ej., mejores métodos diagnósticos que permiten un diagnóstico más precoz, mejores técnicas quirúrgicas...). Los pacientes de la actualidad mejorarán su pronóstico no sólo por la utilización del fármaco experimental, sino por todas esas "cointervenciones". Así, los estudios con controles históricos tienden a **sobreestimar el efecto del fármaco experimental** (MIR).
- Puede existir un problema de homogeneidad entre el grupo experimental y el control si los criterios diagnósticos o para tratar la enfermedad han variado.

Debido a estas limitaciones, la utilización de controles históricos se restringe a situaciones en las que es **muy difícil reclutar el tamaño muestral necesario en el presente** (MIR) (enfermedades raras o terminales).

Además, se exige que las variables resultado sean **variables duras**.

7.8. Realización de muchas comparaciones en los estudios epidemiológicos

La realización de muchas comparaciones en un estudio supone un problema ya que, con cada comparación realizada, existirá una probabilidad de cometer un error alfa (encontrar diferencias que en realidad no existen), y la probabilidad alfa de cada comparación se irá acumulando hasta tener una probabilidad global de haber cometido errores en el estudio muy alta (MIR).

Así, pese a tener comparaciones individuales estadísticamente significativas (si p <0,05), puede ocurrir que no podamos decir que de forma global nuestro estudio ha demostrado encontrar diferencias (p "global" >0,05).

Ejemplo: si realizamos dos comparaciones con un error alfa de 0.04 en cada una, la probabilidad global de haber cometido un error alfa en el estudio será = 0,04 + 0,04 – 0,04 · 0,04 = 0,0784 (resultado "global" no significativo).

Para evitar este problema, se aconseja aplicar una **penalización estadística** (MIR 18, 209): se exigen niveles de significación de cada comparación individual lo suficientemente bajos para que, al sumarlos, el valor alfa "global" sea p <0,05. *De forma aproximada, el nivel de significación exigido para cada comparación individual es pi = 0,05 / n.º de comparaciones.*

Hay varias situaciones en las que se realizan muchas comparaciones en los estudios epidemiológicos:

Análisis de comparaciones múltiples

Se comparan muchas variables esperando que al menos en alguna de ellas se encuentren diferencias significativas. Implica como hemos visto un **mayor riesgo de encontrar falsas diferencias**, por lo que sus resultados deben interpretarse con precaución (MIR).

Análisis de subgrupos

Se analizan los resultados obtenidos en subconjuntos de la muestra (p. ej., en ancianos, en diabéticos, en pacientes con insuficiencia renal...).

Puede ser útil para conocer el comportamiento de un fármaco en distintos grupos poblacionales, pero su realización es muy **sensible a los sesgos**, especialmente si los subgrupos no se han previsto desde el inicio del estudio.

En general, el análisis de subgrupos plantea nuevas hipótesis de trabajo, pero no sirve para confirmarlas. *Dichas hipótesis deberán confirmarse en nuevos estudios realizados específicamente en pacientes del subgrupo de interés.*

Análisis intermedios (MIR)

Son análisis de los resultados que se realizan en momentos intermedios del seguimiento de un estudio, cuando dicho **seguimiento es muy largo** (por ello, se realizan sobre todo en los estudios de fase III).

Se realizan para evitar que pasen desapercibidas diferencias importantes entre los grupos en comparación (que pueden haber surgido antes de que finalice el seguimiento del estudio), y por tanto incurrir en problemas éticos (seguir tratando a un grupo de pacientes con un fármaco inferior a otro): en caso de encontrar **diferencias significativas** en un análisis intermedio, **se detiene el estudio**.

La realización de análisis intermedios debe estar planificada en el diseño del estudio antes de comenzarlo.

Además, hay que tener en cuenta que al hacer análisis intermedios se incurre en comparaciones múltiples, por lo que existe un riesgo de **sobrestimar** el beneficio del tratamiento experimental (mayor probabilidad de cometer un error alfa) (MIR 15, 193). Por ello, habrá que aplicar una penalización estadística.

7.9. Estudios de bioequivalencia

Tipos de especialidades farmacéuticas

- **Fármaco original o innovador**: investigación y desarrollo completo por parte de la industria farmacéutica. Tiene un periodo de exclusividad en el que sólo lo pude comercializar la compañía farmacéutica que lo ha desarrollado.
- **Licencias o segundas marcas**: la compañía farmacéutica que ha desarrollado un fármaco (y todavía lo comercializa en periodo de exclusividad) autoriza a otra empresa para que también lo distribuya.

- **Especialidad farmacéutica genérica (EFG)** (MIR): especialidad con la misma forma farmacéutica (comprimidos, cápsulas, vial para inyección...) e igual composición cualitativa y cuantitativa de principio activo que otra especialidad de referencia, cuyo perfil de eficacia y seguridad esté suficientemente demostrado. Los excipientes pueden ser distintos. Pueden registrarse antes de que haya expirado la patente original, pero se comercializan una vez haya expirado. En el registro han de aparecer las mismas indicaciones de la especialidad original.

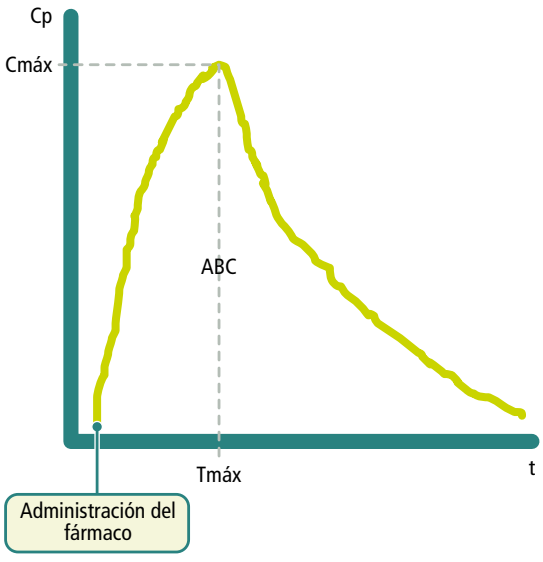
Estudios de bioequivalencia

Para poder registrar un fármaco genérico, éste debe demostrar su equivalencia terapéutica con la especialidad de referencia mediante los correspondientes **estudios de bioequivalencia**.

En ellos, no es necesario demostrar la eficacia y la relación beneficio/riesgo del producto (MIR), siendo suficiente demostrar que las **características farmacocinéticas** (MIR 16, 38; MIR 15, 208) del genérico (concentración plasmática máxima alcanzada, tiempo que se tarda hasta alcanzar esa concentración, y cantidad total de fármaco absorbida –área bajo la curva–) no son significativamente distintas a la del producto original.

Los estudios de bioequivalencia suelen realizarse en voluntarios sanos y con un diseño cruzado (MIR 17, 36). El análisis de los datos se realiza para demostrar **equivalencia terapéutica** con un límite bilateral de un 20% de diferencias. Dichas diferencias se miden en el intervalo de confianza del 90% (IC 90%) del cociente de medias de los parámetros farmacocinéticos mencionados, que debe encontrarse entre el 80-125% (MIR 12, 200). *Se divide la media de los resultados obtenidos en el fármaco original entre la media de los resultados obtenidos con el genérico.*

La demostración de bioequivalencia permite suponer que, ante similar indicación y con la misma pauta posológica, esos productos presentarán la misma eficacia clínica.



Cp: concentración plasmática. Cmáx: concentración plasmática máxima. Tmáx: tiempo que se tarda hasta alcanzar la Cmáx. ABC: área bajo la curva, que es proporcional a la cantidad total de fármaco absorbida.

Figura 8. Parámetros analizados en un estudio de bioequivalencia.

Las diferentes formas farmacéuticas orales de liberación inmediata (comprimidos, cápsulas...) podrán considerarse la misma forma farmacéutica siempre que hayan demostrado su bioequivalencia.

7.10. Estudios farmacoeconómicos

Los estudios farmacoeconómicos permiten analizar los resultados de un fármaco en función de su eficacia y también sus costes; esto es conocido como eficiencia.

Estos análisis pueden ser de dos tipos.

ANÁLISIS PARCIALES	ANÁLISIS COMPLETOS
Descripción de costes Descripción de consecuencias Descripción de costes y consecuencias Evaluación de eficacia y efectividad Análisis de costes (MIR 11, 193)	Minimización de costes Análisis coste-efectividad Análisis coste-utilidad Análisis coste-beneficio

Tabla 3. Tipos de análisis.

Los análisis más importantes son los completos:

Minimización de costes

Cuantifica los costes de dos o más procedimientos cuyas consecuencias son equivalentes (MIR 17, 126; MIR).

Análisis coste-efectividad

Compara el coste que supone en condiciones reales obtener un año de vida ganado, una curación, una muerte evitada, etc.

Para calcular si la mayor efectividad de un fármaco es rentable respecto a otro, realizamos los estudios de **análisis incremental** (MIR 10, 198; MIR), que calculan el ratio de coste-efectividad incremental (MIR 15, 188). Este indicador nos proporciona información sobre si los costes adicionales, originados por un cambio del tratamiento A al tratamiento B, pueden ser justificados por los beneficios clínicos adicionales obtenidos. Si los costes adicionales originados no superan el umbral de coste-efectividad permitido (coste máximo que asumiríamos por cada unidad de efectividad ganada), podemos decir que el fármaco B es coste-efectivo respecto al otro. En cambio, si se supera el umbral de coste-efectividad permitido, el fármaco B no será coste-efectivo y no podremos permitir usarlo.

$$\textbf{RCEI} = \frac{\text{(CM fármaco B - CM fármaco A)}}{\text{(EM fármaco B - EM fármaco A)}}$$

RCEI = Ratio de coste-efectividad incremental entre los fármacos A y B
CM = Coste medio
EM = Efectividad media

Figura 9. Ratio de coste-efectividad incremental.

Análisis coste-utilidad (MIR 19, 124)

Analiza la cantidad y calidad de vida. La utilidad es un concepto estadístico que combina la probabilidad de un desenlace o resultado con las preferencias del paciente respecto a este desenlace (muerte, curación, secuelas...). Una de las medidas de utilidad más empleadas son los **AVAC o QALY**, años de vida ajustados por calidad de vida (MIR 19, 20; MIR 18, 217; MIR 14, 208; MIR 12, 195). La calidad de vida relacionada con la salud se evalúa mediante cuestionarios genéricos o específicos estandarizados.

Otra medida utilizada en este tipo de estudios son los **años de vida ajustados por discapacidad** (**AVAD o DALY** en inglés -*disability adjusted life years*-). Esta unidad comprende los años en los que se acorta la esperanza de vida por la enfermedad (años potenciales de vida perdidos: APVP o YLL en inglés -*years of life lost*-) sumados a los años en los que el paciente vivirá con algún tipo de discapacidad fruto de la enfermedad (años vividos con discapacidad: AVD o YLD en inglés -*years lived with disability*-). Es la medida que mejor valora la carga global de una enfermedad (MIR 22, 49; MIR 20, 178).

Al igual que sucede con los estudios de análisis coste-efectividad, para calcular si la mayor "utilidad" de un fármaco es rentable con respecto a otro, realizaremos los estudios de análisis incremental que calculan el ratio de coste-utilidad incremental (MIR 17, 125). Este indicador nos proporciona información sobre si los costes adicionales, originados por un cambio del tratamiento A al tratamiento B, pueden ser justificados por los beneficios clínicos subjetivos adicionales obtenidos.

Análisis coste-beneficio

Mide los costes y los efectos en términos económicos (MIR 13, 192; MIR).

Recuerda...

PISTA MÍRICA	TIPO DE ANÁLISIS
Consecuencias similares	Minimización de costes
Años de vida ganados	Coste-Efectividad
Años de vida+ Calidad de vida (AVAC)	Coste-Utilidad
Unidades monetarias (Euros) Valora hacer algo- No hacer nada (MIR)	Coste-Beneficio

(Ver tabla 4)

Análisis de sensibilidad en evaluaciones farmacoeconómicas

Es el estudio del impacto de las variaciones en las variables más importantes y/o con mayor incertidumbre en el resultado final del estudio. Variables con mayor **incertidumbre**: costes más importantes (hospitalización, pruebas diagnósticas caras), efectividad, tasa de descuento.

Los resultados serán robustos cuando las modificaciones en el valor de las variables con mayor incertidumbre tengan poco efecto en los resultados. Las modificaciones en las variables de incertidumbre afectarán al coste y a los resultados, modificando el RCEI. El análisis de sensibilidad estudia cómo se modifica el RCEI modificando las variables de incertidumbre (a mayor robustez, menor alteración del RCEI; a menor robustez mayor alteración del RCEI).

Ejemplo. El fármaco A tiene un coste medio de 5000 euros y produce una efectividad media de 3 años de vida ganados; el fármaco B tiene un coste medio de 6000 euros y produce una efectividad media de 4 años de vida ganados. Con estos datos, el RCEI entre los fármacos A y B es = 1000 euros/año vida ganado. Si modificamos el valor de variables incertidumbre que afecten al coste o la efectividad el RCEI se alterará. Por ejemplo, si el coste del fármaco B subiera a 8000 euros, el RCEI sería = 3000 euros/año de vida ganado.

Tipos de análisis de sensibilidad (MIR 18, 216)

- **Simple**: puede ser univariante o multivariante, en función de si se modifica el valor de una o más variables.
- **Análisis umbral**: identifica el valor umbral de la variable por encima del cual se modifican los resultados.
- **Análisis de extremos**: se conocen los resultados al incluir los valores más favorables y más desfavorables de las variables de interés.
- **Análisis probabilístico**: a cada variable se le otorga una distribución de probabilidad, realizando muchas simulaciones y obteniéndose una distribución media de valores. P. ej., método Montecarlo.

Clasificación de los costes

Costes tangibles

- **Directos**:
  - **No sanitarios**: aquellos que inciden sobre pacientes o enfermos, pero que no implican factores o recursos sanitarios. Ejemplo: apoyo social, adaptaciones en el hogar, desplazamiento para buscar atención, etc.
  - **Sanitarios** (MIR 13, 193; MIR 10, 197): aquellos que representan factores o productos sanitarios que son utilizados, consumidos o desgastados. Ejemplo: consumo de fármacos o material sanitario, salarios del personal sanitario, etc.
  - **Negativos**: aquellos que representan ahorros en los recursos sanitarios. Ejemplos: ahorro en servicios e intervenciones evitadas, tratamientos sustituidos, etc.
- **Indirectos**: aquellos derivados de la reducción de la capacidad para generar ingresos, la disminución del rendimiento laboral o del aumento de los costes empresariales (MIR 15, 187; MIR). Ejemplos: tiempo laboral perdido, productividad reducida...

Costes intangibles

Aquellos no valorables por los mecanismos de precio del mercado. Ejemplos: miedo, dolor, incomodidad, ansiedad, molestias, ocio perdido, etc.

	U. DE COSTES	U. DE RESULTADOS	VENTAJAS	INCONVENIENTES
MINIMIZACIÓN DE COSTES	EUROS	Iguales	Sencillo de realizar	No suelen existir consecuencias equivalentes
COSTE-EFECTIVIDAD	EUROS	Físicas (años de vida)	Resultados son fáciles de entender	No pueden compararse programas con diferentes unidades de resultados
COSTE-UTILIDAD	EUROS	AVAC	Valora enfermedades crónicas	Es valoración subjetiva de resultados
COSTE-BENEFICIO	EUROS	EUROS	Compara programas con diferentes unidades de resultados	Difícil transformación de unidades físicas a monetarias

Tabla 4. Tipos de análisis de evaluación y sus características.

Perspectiva de los estudios farmacoeconómicos

En los estudios farmacoeconómicos la perspectiva va a definir el punto de vista desde donde se realiza el estudio. Dependiendo de la perspectiva escogida será necesario incluir unos costes u otros. Siempre que sea posible debería elegirse la perspectiva de la sociedad en global, ya que es la que incluye todo tipo de costes **(MIR 16, 195)**.

- **Perspectiva hospitalaria**: incluye costes directos hospitalarios (gastos de médicos, investigadores, pacientes, farmacéuticos, dirección hospitalaria).
- **Perspectiva extrahospitalaria**: incluye costes directos extrahospitalarios (ayudas de hogar, traslados en ambulancias...) y costes de administración y aseguradoras.
- **Perspectiva de la sociedad**: incluye tanto costes directos como indirectos (falta de productividad del paciente enfermo).

Tema 8

Errores en los estudios epidemiológicos

Autores: *Víctor Rodríguez Domínguez (2), Eduardo Franco Díez (5), Carlos Corrales Benítez (16).*

ENFOQUE MIR

Tema con importancia intermedia. Últimamente lo más preguntado es el sesgo por factor de confusión y los sesgos propios de los estudios de validación de pruebas diagnósticas, pero todos los sesgos han sido preguntados y por tanto debes dominarlos.

Validez y reproducibilidad en los estudios epidemiológicos

Al igual que en un estudio de validación de un nuevo test diagnóstico, en cualquier estudio epidemiológico se tiene en cuenta su validez y reproducibilidad:

Validez (exactitud) (MIR)

Grado en que un estudio mide lo que realmente tenía como objetivo medir.

- La **validez interna** hace referencia a la exactitud de los resultados para los pacientes del estudio (que los resultados sean aplicables a la muestra), y depende de que el diseño del estudio sea correcto (ausencia de **sesgos**) **(MIR 23, 49; MIR 12, 181)**.
- La **validez externa** hace referencia a que los resultados sean aplicables a la población diana **(MIR)**, y depende de lo representativa que sea la muestra de la población (ausencia de **errores aleatorios**) **(MIR 11, 177)**. La validez interna es un prerrequisito de la validez externa.

Reproducibilidad (fiabilidad, precisión)

Grado de un estudio de obtener el mismo resultado si se repitiera en otras muestras distintas en las mismas condiciones.

Depende de lo representativa que sea la primera muestra respecto a la población de la que se obtienen las siguientes muestras para repetir el estudio (por tanto, de la presencia de errores aleatorios) **(MIR)**.

8.1. Errores aleatorios

El hecho de que estudiemos muestras de individuos y no a la población completa puede hacer que nuestra **muestra no sea representativa de la población**.

- **Tipos de errores aleatorios**: error de tipo I (alfa) y error de tipo II (beta) **(ver tema 3.1. Errores en contraste de hipótesis)**.
- **Consecuencia**: afectan a la **validez externa** del estudio (no afectan a la validez interna).
- **Solución**: el **aumento del tamaño muestral** disminuye el riesgo de cometer errores aleatorios y aumenta la potencia estadística del estudio.

8.2. Errores sistemáticos (sesgos)

Son errores debidos a un **diseño inadecuado** del estudio.

- **Consecuencia**: afectan a la **validez interna** del estudio (por lo que, secundariamente, afectan también a la validez externa).
- No se ven influidos por el tamaño muestral **(MIR 23, 44; MIR)**.

Sesgos de selección

Aparece cuando existen **diferencias en las características** que tienen los distintos grupos en estudio (aparte de la característica estudiada), y dichas diferencias influyen en los resultados **(MIR)**. *Por ejemplo, un grupo es más anciano que el otro, lo que influye en que se mueran más sujetos en dicho grupo.*

Siempre que realizamos un estudio comparando varios grupos, debemos confirmar al inicio del análisis de resultados que las características de los grupos sean homogéneas (grupos "comparables" entre sí **(MIR 22, 45; MIR 11, 185)**), en cuyo caso no habrá posibilidad de que existan sesgos de selección.

- **Solución**: el **muestreo aleatorio** (en estudios observacionales) y la **aleatorización** **(MIR 14, 197)** (en estudios experimentales) disminuyen las probabilidades de que las características de los pacientes se distribuyan de forma dishomogénea entre los grupos. Disminuyen las probabilidades de incurrir en un sesgo de selección, aunque como son técnicas que dependen del azar no garantizan la eliminación del sesgo.

  El **análisis de subgrupos**, el **análisis estratificado** y el **análisis multivariante** permiten, a posteriori (tras finalizar el estudio), comprobar si las diferencias en las características de los grupos participantes influyen en los resultados **(MIR)** (y por tanto si suponen un sesgo de selección), pero **no permiten eliminar el sesgo**.

Ejemplos de sesgos de selección

- **Sesgo de autoselección (del voluntario)**: cuando se recluta a pacientes voluntarios para participar en un estudio, suelen ser pacientes que no encuentran alivio con los tratamientos disponibles y buscan "a la desesperada" una solución. Dichos pacientes suelen por tanto estar **más graves** que la media, lo cual puede afectar a los resultados (cualquier fármaco tenderá a ser menos efectivo).
- **Sesgo del obrero sano**: si para estudiar una **enfermedad laboral** se acude al lugar de trabajo para seleccionar a los individuos, se infraestimará la frecuencia de enfermedad, ya que aquellos sujetos enfermos no estarán trabajando sino de baja.

- **Sesgo diagnóstico (de Berkson)**: sesgo que ocurre cuando se seleccionan los individuos de un estudio de entre **pacientes hospitalizados**, y el factor que se está estudiando es un factor de riesgo para hospitalizarse.

  *Ejemplo: se analiza la posible relación causal entre el VIH y los linfomas en pacientes hospitalizados mediante un estudio de casos (pacientes con linfoma) y controles (hospitalizados por otras causas), analizando la frecuencia de VIH en cada grupo. Tanto el VIH como los linfomas son por sí mismos factores de riesgo de ingresar en el hospital (si un individuo con VIH o con linfoma enferma por otras causas, es probable que le ingresen porque su enfermedad de base le convierte en un paciente de alto riesgo). Así, entre los pacientes "hospitalizados por otras causas" habrá un subgrupo de pacientes cuyo motivo de hospitalización será VIH (mientras que el grupo de pacientes con linfoma ya tienen bastante con su tumor para ingresar). Esto hará que **infraestimemos** la asociación entre linfoma y VIH, al encontrar más pacientes VIH en el grupo control.*

- **Sesgo de incidencia/prevalencia (falacia de Neyman)** (MIR): sesgo que ocurre en los estudios de **casos y controles** al estudiar enfermedades que tienen una fase aguda con altas tasas de letalidad y una fase crónica posterior *(p. ej., IAM, ictus, disección de aorta)*, ya que sólo podremos estudiar a los casos "prevalentes" (crónicos que han sobrevivido a la fase aguda) mientras que se nos pasarán desapercibidos los casos "incidentes" (agudos que fallecieron). Las características de los casos prevalentes pueden ser distintas a las de los casos incidentes y eso tener implicaciones en los resultados.

  Ejemplo: al analizar la relación causal de la FA con el ictus, infraestimaremos la asociación porque los ictus cuya causa es cardioembólica por FA tienen una mayor letalidad que los ictus por otras causas. Así, los casos "prevalentes" de ictus tendrán una prevalencia de FA menor que la del conjunto de pacientes con ictus (sumando los casos incidentes y los prevalentes).

Figura 1. Jerzy Neyman, matemático polaco del siglo XX. Entre sus aportes a la Epidemiología (además de describir el sesgo que lleva su nombre) se encuentra el diseño de los estudios de bioequivalencia.

Sesgos de clasificación (de información, de medida)

Aparece cuando se clasifica erróneamente una variable en estudio, pensando que los pacientes que presentan esa variable no la presentan o viceversa. *Por ejemplo, el esfingomanómetro del estudio está estropeado y siempre marca PAS 150 mmHg; clasificaremos mal a los sujetos no hipertensos ya que pensaremos que son hipertensos.*

Sesgo de clasificación incorrecta no diferencial

Aparece por **errores en los aparatos de medida** que condicionan el mismo nivel de error en la clasificación de todos los grupos del estudio.

Dichos errores tienden a diluir el efecto de la exposición o tratamiento estudiados, por lo que **infraestiman** la asociación (MIR 17, 121; MIR).

Su solución pasa por **mejorar los aparatos de medida** (MIR) (aumentar su S y E).

Sesgo de clasificación incorrecta diferencial

Aparece por errores **subjetivos** de los pacientes o del investigador a la hora de clasificar las variables resultado de los pacientes. Estos sesgos suelen aparecer **cuando el paciente o el investigador conoce a qué grupo pertenece el paciente**, de modo que los pacientes que reciban la intervención experimental o los investigadores pueden interpretar una falsa mejoría en **variables subjetivas** (bienestar, dolor, etc.) respecto de los pacientes del grupo control (que reciben o bien nada, o placebo, o un fármaco activo control): la clasificación de las variables es distinta en cada uno de los grupos.

Así, este sesgo **sobreestima** los resultados (MIR 19, 116), ya que el grupo experimental verá sus variables resultado subjetivas artificialmente mejoradas.

Su solución consiste en emplear **técnicas de enmascaramiento (ciego)** (MIR 11, 234):

- **Estudios abiertos** (MIR 12, 188): los pacientes e investigadores conocen qué intervención reciben los pacientes.
- **Estudios ciegos (con enmascaramiento)** (MIR 13, 187; MIR).
  - **Ciego simple**: los pacientes no saben qué intervención reciben (si la experimental o el control).
  - **Doble ciego** (MIR 13, 189): no lo saben ni los pacientes ni los investigadores.
  - **Triple ciego**: no lo saben ni los pacientes, ni los investigadores ni los analistas de los datos (que suelen ser personas independientes de los investigadores).

Dentro de las técnicas de enmascaramiento, la técnica de **doble simulación o "double dummy"** (MIR 12, 190) se utiliza cuando se realiza un estudio en el que se comparan dos tratamientos cuya forma de administración (oral, i.v...), posología (cada 12 h, cada 24 h...) o forma farmacéutica (comprimidos, cápsulas, supositorios...) son distintos.

Dicha técnica consiste en administrar, en cada uno de los grupos del estudio, el fármaco que le toca a dicho grupo, y además el placebo del fármaco que le toca al otro grupo.

Ejemplos del sesgo de clasificación incorrecta diferencial son:

- **Sesgo del entrevistador**: ocurre cuando el investigador no está enmascarado, y dirige o interpreta de manera inconsciente la entrevista con el paciente, de modo que parezca por sus respuestas que ha mejorado.
- **Sesgo de memoria o amnésico** (MIR 14, 198; MIR 10, 186): en los estudios de casos y controles, los casos suelen recordar más la exposición al factor de riesgo que los controles.
- **Sesgo de atención o efecto Hawthorne**: los participantes de un estudio pueden actuar de modo distinto del habitual simplemente por sentirse observados.

  Por ejemplo: un ensayo clínico estudia la eficacia de la rosuvastatina para el tratamiento de la hipercolesterolemia. Todos los pacientes deberán seguir una dieta baja en grasas y realizar ejercicio, y a un grupo de pacientes, además, se les administrará rosuvastatina. Los pacientes que sepan que les ha tocado el fármaco experimental realizarán en mayor proporción ejercicio físico que los que sepan que no reciben tratamiento, ya que tratarán de "ayudar" al fármaco con más ilusión y esperanza para mejorar su salud.

Figura 2. Trabajadoras de la fábrica Hawthorne Works, de la compañía Western Electric. El efecto Hawthorne recibe su nombre por unos estudios sobre productividad industrial realizados en dicha fábrica de la localidad de Cicero (Illinois) entre 1924 y 1932. El estudio más famoso consistió en comparar la productividad industrial con iluminación ambiental más alta o más baja, y sus resultados fueron que la productividad aumentó tanto en el grupo sometido a alta iluminación, como en el sometido a baja iluminación; fue el hecho de saberse observados el que propició que las trabajadoras aumentaran su productividad.

Sesgo por factor de confusión (FC)

Un factor de confusión es un **factor de riesgo** para la enfermedad en estudio, y que además **se asocia estadísticamente a la exposición** cuya asociación causal con la enfermedad queremos estudiar (MIR 14, 200; MIR). Además, el factor de confusión debe **actuar de forma independiente** a la exposición en cuanto al mecanismo por el que provoca la enfermedad (no puede ser un paso intermedio en la relación exposición-enfermedad).

Como el FC y la exposición se asocian estadísticamente, los pacientes del grupo expuesto presentarán en mayor proporción el FC que los pacientes del grupo no expuesto, independientemente del método de selección de la muestra (aunque sea aleatorizada). Así, al suponer el FC un riesgo para la aparición de enfermedad, **parte del riesgo que atribuyamos a la exposición se deberá realmente al FC** (**sobreestimando** pues la verdadera asociación causal).

El sesgo por factor de confusión es el **único que puede eliminarse a *posteriori*** (MIR 12, 177):

Soluciones a priori del sesgo por factor de confusión (MIR 18, 210; MIR)

- **Restricción**: el FC supone un criterio de exclusión para el estudio. Así, ningún paciente (ni del grupo expuesto ni del no expuesto) tendrá el FC y éste no podrá influir en los resultados.
- **Apareamiento** (MIR 11, 178; MIR): por cada paciente incluido en el grupo expuesto que posea el FC, incluiremos un paciente en el grupo no expuesto que lo posea. Así, el porcentaje de pacientes con el FC será el mismo en ambos grupos, y se eliminará la influencia del FC sobre la asociación exposición-enfermedad.

En los estudios experimentales, la **aleatorización** disminuye la posibilidad de cometer un sesgo por factor de confusión, pero no es un método que garantice la eliminación del sesgo, dado que depende del azar.

Soluciones a posteriori del sesgo por factor de confusión (MIR)

- **Análisis de subgrupos y análisis estratificado**: mediante estos análisis estadísticos podemos calcular exactamente qué porcentaje del riesgo inicial era atribuible al FC, y por tanto eliminar dicho riesgo para quedarnos con el riesgo atribuible únicamente a la exposición estudiada.
- **Análisis multivariante**: si incluimos como variables independientes (x_i) de una regresión múltiple tanto a la exposición como al FC, el coeficiente de cada una de estas variables quedará "ajustado entre sí", indicando exclusivamente el riesgo atribuible a cada variable individual.

Recuerda...

El análisis de subgrupos y el análisis estratificado permiten, a posteriori, detectar el sesgo de selección (pero no eliminarlo), así como detectar y eliminar el sesgo por factor de confusión.

Tanto el sesgo de selección como el sesgo por factor de confusión se sospechan cuando existen diferencias significativas en alguna característica de los grupos en comparación. Los resultados del análisis de subgrupos o estratificado que realicemos permitirán saber si estamos ante un factor de confusión o un sesgo de selección.

El resultado del sesgo de selección se denomina **factor modificador del efecto**, y actúa (al contrario que el factor de confusión) de forma dependiente a la exposición (potenciando su efecto). Puede ser además un factor de riesgo independiente de la enfermedad, pero no tiene por qué serlo.

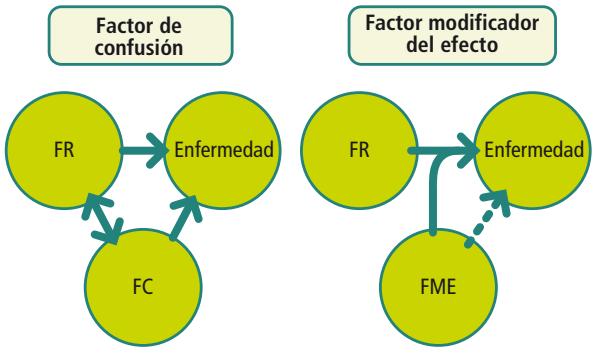


Figura 3. Factor de confusión y factor modificador del efecto.

Sesgo de atricción

En cualquier estudio epidemiológico puede haber pérdidas, y dichas pérdidas pueden ser de dos tipos: las pérdidas **pre-aleatorización** se producen cuando los pacientes no cumplen los criterios de selección del estudio, y afectan por tanto a su validez externa. Las pérdidas **post-aleatorización** se producen en los estudios prospectivos tras la asignación del tratamiento, y pueden afectar a la validez interna.

Cuando existen **diferencias en el porcentaje de pérdidas post-aleatorización** de los distintos grupos de un estudio prospectivo, y las pérdidas no se incluyen en el análisis estadístico de los resultados, aparece un sesgo de atricción **(MIR)**.

Habitualmente, si los dos grupos tienen características homogéneas, la diferencia en las pérdidas se deberá a un efecto adverso del tratamiento experimental por el cual los sujetos de dicho grupo abandonan más el estudio que los del otro grupo.

La solución del sesgo de atricción consiste en estudiar los resultados mediante un **análisis por intención de tratar** en lugar de un análisis por protocolo.

- **Análisis por protocolo (MIR)**: sólo se estudian los resultados de los pacientes que finalizan el estudio. Si existe diferente proporción de pérdidas entre el grupo experimental y el control y estas pérdidas se deben a efectos adversos del fármaco experimental, **sobreestimaremos** el beneficio del fármaco experimental al no tener en cuenta a los pacientes que tienen que dejar de tomarlo (y por tanto dejan de beneficiarse de él).

  Es un peor tipo de análisis y sólo se permite realizarlo en los estudios con diseño de **no inferioridad (MIR)**.

- **Análisis por intención de tratar (MIR 13, 188)**: se estudian los resultados de todos los pacientes reclutados en el estudio (todos los pacientes aleatorizados) **(MIR)**, incluyendo a los pacientes que cursen pérdida o que sean traspasados entre grupos, considerándose cada paciente como perteneciente al grupo al que fue aleatorizado *(p. ej., un paciente aleatorizado a tratamiento médico al que finalmente se somete a cirugía por fracaso del tratamiento médico se considera un fracaso del tratamiento médico y no un éxito de la cirugía)*. Permite estudiar la causa de las pérdidas y el efecto global del fármaco teniendo en cuenta que un porcentaje de pacientes no se lo tomará (situación que simula a la que se observará en la **práctica clínica real** una vez se comercialice el fármaco).

  Es el mejor tipo de análisis y el único permitido en los estudios con diseño de **superioridad**.

8.3. Sesgos específicos de los estudios de validación de pruebas diagnósticas

Existen sesgos específicos cuando intentamos testar la validez de una nueva prueba de diagnóstico, mediante su comparación con la prueba *gold standard* **(ver tema 5. Estudios de validación de una prueba diagnóstica)**.

De modo general, para prevenir estos sesgos se deben cumplir las mismas premisas que para el resto de estudios. De manera particular, en estos estudios es importante lo siguiente:

- El estudio se debe realizar en la población diana a la que va dirigida la nueva prueba diagnóstica.
- A todos los pacientes se les debería realizar la nueva prueba diagnóstica y la prueba *gold standard* (de otro modo no es posible conocer la sensibilidad/especificidad real).
- Los estudios deberían ser siempre enmascarados; no se debe conocer el resultado del gold estándar al realizar el test nuevo, ni viceversa.

Sesgos de selección en estudios de validación de pruebas diagnósticas

Afectan específicamente a estudios de validación de **pruebas de *screening***.

Sesgo de duración de la enfermedad (*length-time bias*) (MIR 20, 33)

Sobreestima el efecto de los tests de *screening* **en enfermedades leves respecto a las graves**. Las enfermedades leves suelen tener una evolución más lenta y por tanto una mayor duración total que las enfermedades graves (cuya evolución suele ser más rápida). Por ello, como los pacientes con enfermedades leves viven más años con la enfermedad, tendrán más posibilidades de ser diagnosticados con pruebas de *screening* realizadas con una determinada periodicidad que los pacientes con enfermedades graves.

De este modo, parece que la realización del test y su diagnóstico precoz supone un aumento en la supervivencia solo en las enfermedades leves.

Este sesgo es mayor a medida que la prueba es **menos sensible**. *Las pruebas poco sensibles detectan peor la enfermedad, tanto los casos leves como los graves. Los casos graves duran menos tiempo y por tanto reciben menos número de tests de screening (menos probabilidades de ser detectados); sin embargo los casos leves duran mucho tiempo y reciben varios tests de screening durante la duración de la enfermedad, por lo que tendrán más probabilidades de ser detectados. Así, se sobreestima aún más la ventaja del screening en las enfermedades leves respecto a las graves.*

(Ver figura 4)

Sesgo de sobrediagnóstico

Si el tiempo de evolución hasta la aparición de síntomas tiende a aproximarse a los años de vida del paciente, una enfermedad (p.ej. cáncer de próstata en un hombre de 100 años) detectada por una prueba de *screening* supone

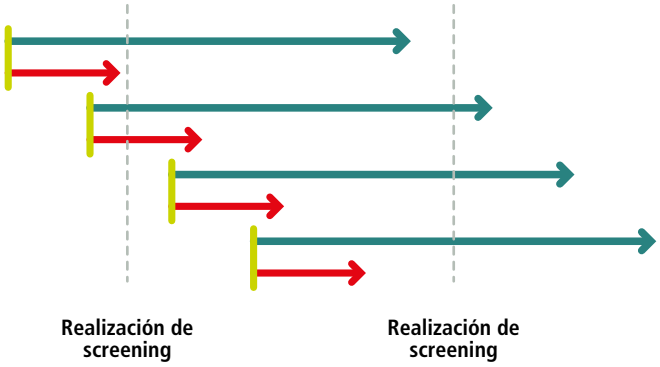


Figura 4. Sesgo de duración. Representamos a 4 sujetos con una enfermedad, que puede tener curso leve (en verde, con larga duración) o curso grave (en rojo, con duración corta de la enfermedad). La realización de *screening* es capaz de detectar a todos los casos leves (incluso al segundo paciente lo detectaríamos en las dos pruebas de *screening*), pero solo a uno de los casos graves

un *sobrediagnóstico*: su detección no aporta beneficios al paciente, ya que iba a fallecer por otra causa sin llegar a presentar síntomas.

Por tanto, se **infraestima** la utilidad del test.

Sesgo del voluntario sano

Habitualmente, los sujetos que se presentan como voluntarios a estudios de pruebas de *screening* son personas más comprometidas con su estado de salud y a cooperar para mejorarla. Esto lleva a que, independientemente del cribado, estos pacientes tengan una expectativa de vida mayor (hábitos de vida más saludables).

El efecto de este sesgo actúa, por tanto, **sobreestimando** el efecto del *screening*.

¡OJO! No confundáis este sesgo con el sesgo de autoselección (del voluntario) en los ensayos clínicos; en dicho caso, los pacientes que suelen prestarse a participar en los ensayos clínicos son los más graves (se infraestima la eficacia del tratamiento experimental).

Sesgos de clasificación en estudios de validación de pruebas diagnósticas

Son sesgos de clasificación incorrecta **diferencial**.

Sesgo de adelanto diagnóstico (*lead-time bias*)

Ocurre cuando realizamos pruebas diagnósticas de ***screening*** a pacientes en fase pre-clínica. En este caso estamos realizando un diagnóstico cuando el paciente está todavía asintomático y, por tanto, la duración confirmada de la enfermedad será mayor que si realizásemos la prueba solamente a pacientes que ya están en fase clínica. Así, se alarga el tiempo que el paciente vive con la enfermedad cuando hemos realizado un test del *screening*, y con ello parece que mejora la supervivencia.

Este sesgo **sobreestima** el efecto que tiene el método de *screening*, pareciendo que aumenta la supervivencia de la enfermedad, cuando en realidad lo que aumenta es el tiempo en que somos conscientes de la presencia de la enfermedad.

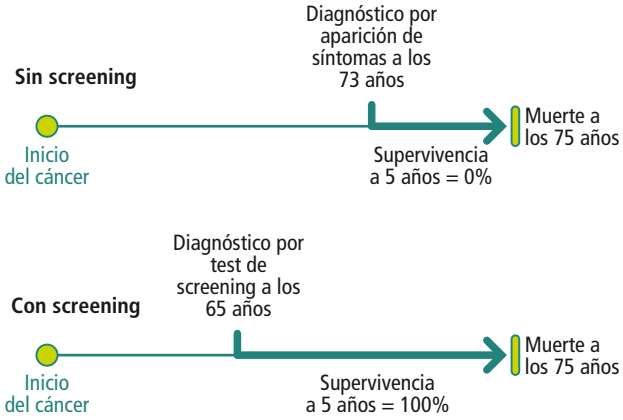


Figura 5. Sesgo de adelanto diagnóstico.

Sesgo de verificación o de validación (MIR 19, 130)

Ocurre en **estudios retrospectivos**. En la práctica clínica habitual, a los pacientes que dan positivo en las pruebas diagnósticas iniciales se les realizan pruebas adicionales (más específicas) para confirmar el diagnóstico, mientras que a los pacientes que dan negativo, en cambio, no se les realizan más pruebas. Esto hace que sea imposible calcular de manera correcta la sensibilidad y especificidad de las pruebas iniciales (no sabemos el total de enfermos ni de sanos, al no haber hecho la prueba *gold standard* en los pacientes negativos); sólo permitirá calcular su **valor predictivo positivo** (en los pacientes que dieron positivo sí hemos realizado el test *gold standard* y sabemos cuántos son enfermos y cuántos sanos).

Si realizamos estudios prospectivos, en cambio, podremos incluir en el protocolo del estudio que se realizará la prueba *gold standard* en todos los pacientes, sean positivos o negativos en el test diagnóstico que queremos validar; esto evitará incurrir en un sesgo de verificación.

Sesgo de sospecha diagnóstica

Sobreestima el efecto del test diagnóstico. Ocurre en estudios **tanto** prospectivos como retrospectivos, cuando **no se enmascara** el resultado del *gold standard* a los evaluadores de los resultados del test diagnóstico que estamos evaluando, o viceversa. Conocer el resultado de una de las pruebas puede influir en la interpretación de la otra prueba realizada sobre el mismo paciente. Este sesgo se corrige **enmascarando** dichos resultados.

Por ejemplo, si un médico debe decidir si una prueba diagnóstica con componente subjetivo (por ejemplo una radiografía de tórax) es positiva o negativa, y conoce que previamente al sujeto se le realizó otra prueba que dio positivo, tendrá mayor tendencia a informar la radiografía de tórax también como positiva en los casos dudosos.

Error	Consecuencia	Solución
Errores aleatorios	↓ Validez externa	↑ n
Errores sistemáticos	↓ Validez interna → ↓ Validez externa	Mejorar diseño
Selección	Habitualmente **sobreestima** * Los ejemplos concretos (voluntario, obrero sano, Berkson, Neyman) suelen infraestimar	Muestreo aleatorio (estudios observacionales) Aleatorización (estudios experimentales)
Clasificación incorrecta no diferencial	**Infraestima** la asociación	Mejorar aparatos de medida
Clasificación incorrecta diferencial	**Sobreestima** la asociación	Enmascaramiento
Factor de confusión	**Sobreestima** la asociación	**A priori:** - restricción - apareamiento - aleatorización (estudios experimentales) **A posteriori:** - análisis de subgrupos - análisis estratificado - análisis multivariante
Atricción	**Sobreestima** la asociación	Análisis por intención de tratar

Figura 6. Errores en los estudios epidemiológicos.

Valores normales en Estadística y Epidemiología

CONCEPTO	VALORES NORMALES
Significación estadística (error α)	$p < 0,05$
Error β	$\beta < 0,2$
Potencia	Potencia >0,8
Intervalo 68% (de confianza en est. inferencial)	$\mu \pm \sigma$ (eem en inferencial)
Intervalo 95% (de confianza en est. inferencial)	$\mu \pm 2\sigma$ (2eem en inferencial)
Intervalo 99% (de confianza en est. inferencial)	$\mu \pm 2,5\sigma$ (2,5eem en inferencial)
EEM	$EEM = \sigma / \sqrt{n}$
NNT	NNT = 100 / RAR
Límite de no inferioridad (δ = delta)	Habitualmente δ = 20%

Tabla 1. Valores normales en Estadística y Epidemiología.

Reglas mnemotécnicas
Estadística y Epidemiología

Regla mnemotécnica
Para recordar el error alfa: α-fetoproteína (**α-FP**). El error tipo α es un resultado falso positivo (**FP**).

Regla mnemotécnica
Test de contraste de hipótesis para variables cualitativas ***CHI*** *tuviera un* ***YATE*** *iría a* ***PESCAR*** *a* ***NEMO*** Datos independientes: **CHI** cuadrado Corrección de **YATE**s Corrección de **Fisher (PESCAR)** Datos apareados: Test de Mc**NEMAR**

Bibliografía

Argimón Pallás, J. M., Jiménez Villa, J., (2018). *Métodos de investigación clínica y epidemiológica, (4.ª ed.)* Madrid:Elsevier.

Carrasco de la Peña, J. L., (1995). *El método estadístico en la investigación médica.* (6.ª ed.). Ciencia 3 Distribución.

Hernández-Aguado, I., Lumbreras Lacarra, B. (2018). *Manual de Epidemiología y Salud Pública para grados en ciencias de la salud, (3.ª ed.)* Madrid: Editorial Médica Panamericana.

Notas

Sedes

AMIR


Santiago
Oviedo
Bilbao
Pamplona
Valladolid
Zaragoza
Barcelona
Salamanca
Madrid
Tarragona
Valencia
Albacete
Alicante
Córdoba
Sevilla
Murcia
Granada
Málaga

FILIACIÓN PROFESIONAL DE AUTORES

(1) H. G. U. Gregorio Marañón. Madrid.
(2) H. U. La Paz. Madrid.
(3) H. U. Severo Ochoa. Madrid.
(4) H. U. Virgen del Rocío. Sevilla.
(5) H. U. Ramón y Cajal. Madrid.
(6) Le Bonheur Children's Hospital. Memphis, Tennessee, EE.UU.
(7) H. Infanta Cristina. Parla, Madrid.
(8) H. C. San Carlos. Madrid.
(9) H. U. 12 de Octubre. Madrid.
(10) H. U. Fundación Alcorcón. Alcorcón, Madrid.
(11) H. Clínic. Barcelona.
(12) H. U. de Fuenlabrada. Fuenlabrada, Madrid.
(13) H. U. i Politècnic La Fe. Valencia.
(14) H. Vithas Nuestra Señora de América. Madrid.
(15) H. U. de Getafe. Getafe, Madrid.
(16) H. U. Fundación Jiménez Díaz. Madrid.
(17) H. U. Doctor Peset. Valencia.
(18) H. U. Rey Juan Carlos. Móstoles, Madrid.
(19) H. U. Puerta de Hierro. Majadahonda, Madrid.
(20) H. Can Misses. Ibiza.
(21) Psiquiatra en ámbito privado. Madrid.
(22) Centre d'Ophtalmologie Sainte Odile. Alsacia, Francia.
(23) H. Quirónsalud A Coruña. La Coruña.
(24) H. U. Infanta Sofía. San Sebastián de los Reyes, Madrid.
(25) H. U. San Juan de Alicante. Alicante.
(26) H. Moisès Broggi. Sant Joan Despí, Barcelona.
(27) H. U. de Basurto. Bilbao.
(28) H. Virgen de los Lirios. Alcoy, Alicante.
(29) H. U. Central de Asturias. Oviedo.
(30) H. U. Vall d'Hebron. Barcelona.
(31) H. de la Santa Creu i Sant Pau. Barcelona.
(32) H. U. Infanta Elena. Valdemoro, Madrid.
(33) Salud Mental Brians 2, PSSJD. Barcelona.
(34) H. C. U. Lozano Blesa. Zaragoza.
(35) H. U. La Princesa. Madrid.
(36) H. G. U. de Alicante. Alicante.
(37) Directora Academic & Innovation, AMIR.
(38) H. U. de Torrejón, Torrejón de Ardoz, Madrid y H. HM Puerta del Sur, Móstoles, Madrid.
(39) H. C. U. de Valladolid. Valladolid.
(40) H. Infantil U. Niño Jesús. Madrid.
(41) Children's Hospital of Philadelphia. Philadelphia, Pensilvania, EEUU.
(42) H. U. de Torrejón. Torrejón de Ardoz, Madrid.
(43) Instituto Médico y Quirúrgico del Aparato Digestivo IMEQ. Madrid.
(44) H. U. La Princesa. Madrid.
(45) H. U. Infanta Elena. Valdemoro, Madrid.
(46) H. U. del Henares. Coslada, Madrid.
(47) H. G. U. Gregorio Marañón, Madrid y Hospital HM Puerta del Sur, Móstoles, Madrid.
(48) H. U. Principe de Asturias Alcalá de Henares. Madrid.
(49) H. Clínico Universitario de Valencia. Valencia.
(50) H. U. Fundación Jiménez Díaz. Madrid.
(51) H. U. HM Sanchinarro. Madrid.
(52) Médico Especialista en Aparato Digestivo. Madrid.
(53) Instituto de Salud Global de Barcelona (ISGLOBAL). Barcelona.
(54) H. U. San Pedro. Logroño, La Rioja.

La opción correcta

AMIR

www.academiamir.com

9 788419 895264