

A decorative graphic on the right side of the page. It features three concentric blue circles of varying sizes. Two thin blue lines intersect at a point between the top and middle circles, extending towards the top-left and bottom-right corners of the page. A third thin blue line extends from the bottom-right corner towards the bottom-right circle.

Anexo: Planteamiento, estudio y
clasificación de los diseños de
investigación.

Juan José Granizo Martinez



TEMA 1. PLANTEAMIENTO BÁSICO DEL DISEÑO DE INVESTIGACIÓN

1.1. El método científico en Ciencias de la Salud.

Decía Lord Kelvin – un eminente ingeniero que hizo grandes aportaciones a la física, entre ellas la creación de una escala absoluta de temperaturas que mide en grados Kelvin- que la Ciencia se basa en medir y comparar: todo lo demás es opinión.

Esta frase es la mejor definición de la Ciencia: la sustitución de la elucubración y la abstracción por un cuerpo organizado de pensamiento destinado a generar hipótesis en base a la observación objetiva.

El método científico empieza por la observación controlada de un hecho.

Controlada quiere decir que requiere un sistema reconocido, comprobable y reproducible de observación. Estas observaciones se repiten hasta lograr unos resultados fiables, pero en ocasiones, el coste o la complejidad de la medición limitan el número de las mismas. También las limitaciones éticas pueden hacerlo.

De la repetición de las observaciones se infiere (por un proceso intelectual llamado INDUCCIÓN) una norma de comportamiento natural o ley científica (pongamos por caso la Ley de Boyle: si un gas tiene una presión y volumen, si reduzco el volumen debe aumentar la presión. Esta ley posibilita que funcione un motor de explosión.

Introducimos un volumen de gas en un cilindro comprimible: si bajamos el émbolo del cilindro, se reduce el volumen y en consecuencia aumenta la presión. De ello, inferimos que la ley de Boyle se formula de una manera matemática: el producto de presión por volumen al inicio es igual al producto de presión por volumen al final.

$$P_1 \times V_1 = P_2 \times V_2$$



Donde P es presión , V es volumen en los momentos inicial (1) y final (2).

Por cierto, las leyes científicas se expresan como ecuaciones matemáticas, lo que dice mucho de la naturaleza predecible de los sucesos en el mundo de la física clásica...

Una vez que tenemos formulado esta ley, aplicamos este conocimiento deduciendo lo que ocurrirá en nuevos experimentos.

Esta es la fase DEDUCTIVA o de comprobación. Si efectivamente, lo observado coincide con la predicción, la ley científica es válida. Si observamos diferencias, tendremos que concluir que la ley es incorrecta o que su aplicación solo es válida en unas ciertas condiciones.

En ese caso, el proceso se vuelve a poner en marcha para redefinir, corregir o ampliar la ley descrita previamente.

1.2. ¿Por qué el método científico no es aplicable en ciencias de la salud?

Lo que hemos dicho hasta ahora ha funcionado muy bien en la física clásica (la que definió Newton y los padres de la termodinámica....) pero con la física cuántica empezó a tener algunos problemas, aunque en general, las matemáticas siguen siendo el motor del avance en física.

La observación controlada en Ciencia se denomina experimento. El experimento básico consiste en comparar observaciones que en principio son **exactamente iguales**, introduciendo en una de ellas un cambio y comparando el resultado.

Por ejemplo, tenemos un volumen de gas a P y T^a conocidas, provocamos un cambio en ese modelo y medimos como ese cambio modifica las variables.

Si el investigador aumenta la Temperatura secundariamente aumenta la presión del gas.



En otra experiencia con exactamente la misma cantidad de gas a la misma presión y temperatura, lo que hacemos es aumentar la presión, reduciendo el volumen y trascurrido un mismo tiempo, vemos como aumenta la temperatura.

Esta ley física se expresa como $P \times V = n \times R \times T$

Donde P es presión, V es volumen, n es la cantidad en moles del gas, R es una constante y T es la temperatura.

El problema radica en que en física o química es posible repetir de manera exacta las condiciones del experimento (1 gr de nitrógeno, 1 atm. de presión, 25 °C, 2 L de volumen), de manera que si modificamos solo una de esas condiciones, los cambios que se observen serán debidos con certeza a la modificación introducida por que sabemos que nada más ha cambiado.

En ciencias de la salud, esto no es posible. De ahí que el método científico no sea aplicable directamente. En biología no se manejan modelos replicables, sencillamente porque los animales, las personas o las células que manejamos no son nunca exactamente iguales y además cambian con el tiempo. Existe una **variabilidad** que hace imposible el concepto de experimentación tradicional que aplican los físicos.

Y aunque sabemos que no son iguales dos células, solo conocemos algunas de las cosas que las hacen diferentes, a decir verdad, tan pocas que es científicamente imposible decir que dos células, dos personas, dos muestras de tejido, aun siendo de la misma especie, son dos modelos iguales.

Por tanto, habrá respuestas diferentes porque diferentes son las personas, diferentes son sus células.

En ciencias de la salud todavía desconocemos muchas cosas que pueden resultar decisivas en el fin del “inferir” una ley científica y con ella predecir (“deducir”) el resultado. Lo que el investigador obtiene de sus datos es una información confusa: unos pacientes reducen su tensión arterial con un



medicamento: unos mucho, otros poco. Algunos nada. E incluso unos pocos, consiguen lo contrario de lo que se buscaba.

La mayoría toleran bien el medicamento, pero unos pocos no, pero cada uno presenta unos síntomas distintos. En definitiva, aparece VARIABILIDAD.

Esto ya ocurría en física, pero de un modo limitado. Aunque en ocasiones la física también tiene problemas similares, como ocurre cuando se intenta predecir el lugar de impacto de un disparo de cañón. Cada disparo cae en un sitio más o menos agrupado alrededor del lugar teórico de impacto. Esta VARIABILIDAD es uno de los problemas principales en ciencias de la salud.

Pero hay más limitaciones para aplicar el método científico sin más en ciencias de la salud: En este campo, no podemos hacer muchos experimentos por motivos éticos y lo más que alcanzamos es a trasladar los conocimientos adquiridos en animales a humanos lo que no siempre es del todo correcto.

La imposibilidad ética y material de hacer ciertos experimentos obliga a buscar sustitutos al método científico tradicional.

Vamos a ir un poco más adelante. En ciencias de salud no todos los problemas se pueden manejar con una aproximación biológica. En medicina y sobre todo en enfermería hay temas que no aceptan esta aproximación: por ejemplo la cumplimentación terapéutica (¿Por qué unos pacientes cumplen mejor el tratamiento que otros?), la frecuentación de servicios sanitarios (¿Por qué unos enfermos emplean más los recursos sanitarios que otros con la misma patología y gravedad?) solo por poner dos ejemplos.

Estos problemas, que están muy cerca de las ciencias sociales (Ciencias que no son de la naturaleza si no que buscan comprender comportamientos y actitudes) padecen el mismo problema metodológico que las ciencias de la salud.

Para todos estos casos recurrimos a observar que hacen los pacientes y medir los cambios en un proceso que no es experimental. El investigador observa lo



que ocurre en la naturaleza, o lo que hacen libremente los pacientes, pero no manipula experimentalmente los sucesos, es decir, no interviene en ellos.

Este “sucedáneo” del experimento es propio de una disciplina llamada epidemiología de la que toma nombre la metodología investigadora en ciencias de la salud: el método epidemiológico.

1.3. El método epidemiológico.

Epidemiología viene de las palabras griegas “epi” (por encima) y “demos” (población, pueblo). Por tanto, es un conocimiento que se basa en analizar el comportamiento de grandes grupos humanos (poblaciones) a partir de la generalización de observaciones individuales.

Es fácil asimilar los pasos del método científico a los del método epidemiológico.

En un primer paso se realizan observaciones de datos individuales (pacientes) y con ellos se realizan inferencias o generalizaciones que son aplicables a poblaciones. Se pasa de lo concreto a lo general.

Esta fase inductiva toma datos de personas (tienen fiebre, dolor de tórax, tos, esputo purulento...) y con ellos define circunstancias o casos generales y abstractos (“neumonía” como diagnóstico de probabilidad de todo sujeto que tenga los síntomas enumerados).

Este sucedáneo de la inferencia es una tarea propia de la epidemiología clásica.

Por tanto la fase inductiva del método epidemiológico corresponde a la tarea de la epidemiología tradicional o clásica.

Luego, por un proceso deductivo, se pasa de nuevo de lo general a lo particular, intentando hacer una predicción de un suceso. A este proceso deductivo se le llama epidemiología clínica, y lo que busca es aplicar el conocimiento adquirido



a través de la observación de una población a un paciente concreto: si el paciente tiene fiebre, dolor de tórax, tos, esputo purulento...padecerá muy probablemente neumonía. Del conocimiento previo deduzco un tratamiento, un pronóstico y unas complicaciones esperables (por ejemplo, sabemos que con Penicilina el 80 % curan en 7 días, sabemos que su mortalidad es del 5 % y que un 1 % desarrollan daño renal).

La fase deductiva es lo que se denomina epidemiología clínica. Trata sobre el pronóstico, complicaciones y el diagnóstico.

Si resultara que mi predicción no es válida, por ejemplo, por que el paciente fallece prematuramente o apareciera una complicación inesperada, debo revisar mi “idea de la neumonía”, es decir, mi hipótesis: o bien no estamos frente a una neumonía, o ésta neumonía presenta rasgos especiales (esto indica una nueva bacteria o virus, posiblemente).

En otras ocasiones, la fase de observación o inferencia consiste en hacer estudios sobre los factores de riesgo de una población (p. ej. La diabetes se asocia con pacientes obesos, el hábito de fumar aumenta la incidencia de cáncer de pulmón...)

En el proceso de deducción (epi. Clínica) se aplican las consecuencias de este conocimiento adquirido a nuevos sujetos (debemos bajar el peso de los pacientes para prevenir la diabetes, evitar la exposición al tabaco, saber cuánto tenemos que bajar el colesterol o la tensión para tener un riesgo cardiovascular aceptable, lo que es fundamental para marcarse objetivos terapéuticos...).

Como todos sabemos, no todo obeso es diabético, ni todo diabético es obeso. Hay una VARIABILIDAD que hace que en ciencias de salud no digamos en rigor que **el factor de riesgo A causa la enfermedad B**.

Anteriormente he empleado de manera deliberada expresiones como “el diagnóstico de probabilidad es la neumonía”. Una de las consecuencias más



devastadoras de la variabilidad es que desaparece la certeza que daban los matemáticos a los físicos.

El método epidemiológico emplea esta terminología: **el factor de riesgo A aumenta las probabilidades de padecer la enfermedad B.**

Por esa razón, A no es una causa de la enfermedad, es un factor de riesgo: A es un factor de riesgo para padecer la enfermedad B en comparación con no tener el factor A, es decir aumenta las probabilidades, pero no supone ningún tipo de certeza.

Por tanto, estudiar unas pocas personas no sirve de nada, ya que es necesario seguir a toda una población para poder inferir que A es un factor de riesgo de B.

La variabilidad implica falta de certeza. Pero en realidad lo que posiblemente traduce es que todavía no somos capaces de entender todas las causas que operan en un mecanismo fisiológico o en producir una enfermedad.

En el momento actual, la Ciencias de la Salud dan solo probabilidad de algo ocurra, por esta razón se emplean modelos matemáticos aleatorios (no deterministas), es decir, métodos estadísticos en lugar del cálculo que emplean los ingenieros o los físicos.

En esos modelos se calculan las probabilidades de que ocurra un suceso (por ejemplo, morir por un infarto) dadas unas condiciones previas (valores de tensión arterial, colesterol...).

Los modelos aleatorios no asumen que las cosas ocurran por casualidad.

Cuidado con confundir estos términos. Los modelos aleatorios, basados en un cálculo de probabilidades se basan en asumir que una parte de las causas reales de la enfermedad son desconocidas y a todas ellas juntas las asignamos un



valor de riesgo “aleatorio” (seleccionado al azar de un conjunto de valores posibles) en función de las características del paciente.

En la medida en que nuestro conocimiento de las causas mejore hemos de asumir que nuestra capacidad de predecir resultados se irá aproximando más a lo que conocemos del mundo de la física.

TEMA 2. DISEÑO DE UN ESTUDIO.

2.1. Diseño de un estudio de investigación en Ciencias de la Salud.

A la vista de lo dicho en el capítulo anterior, un estudio de investigación tiene dos objetivos básicos: medir un hecho y (en algunos casos) comparar.

Cuando decimos medir, hablamos de una descripción en el más amplio sentido del término.

Un estudio epidemiológico es un proceso que consta de varias fases, que por resumir podemos concretar en tres pasos:

- Definición del problema de investigación,
- Elección del diseño del estudio
- Planificación de las actividades relacionadas con la ejecución del estudio (este tema no es propiamente epidemiológico, pero es muy recomendable que el equipo de investigación se pare a pensar todos los detalles relacionados con la ejecución práctica del estudio).



La primera fase – definir el problema que se quiere investigar- consiste en concretar con meridiana exactitud el problema , que se va a investigar, lo que ha de incluir una serie de pasos previos, que tradicionalmente son:

- Justificación del estudio (¿por qué se quiere hacer?).
- Propósito general (¿Qué se quiere hacer?).
- Objetivos específicos (¿Cómo se va a hacer?).
- Definición de la hipótesis operativa, que ha de ser contrastable por métodos epidemiológicos.

Esta fase es crucial. No es justificable investigar lo conocido, ni emplear una metodología incorrecta, pero tampoco plantearse una pregunta que no tiene respuesta con las herramientas de la ciencia moderna ni con los recursos del investigador.

En la segunda fase se decide el diseño que se va a emplear para dar respuesta a los objetivos y contrastar la hipótesis de investigación. No exageramos si decimos que esta decisión es crítica e irreversible, por lo que debe ser cuidada con mimo por el equipo de investigación.

La respuesta a un estudio se encuentra sin lugar a dudas, en el material y métodos empleados por los investigadores.

Esta decisión debe realizarse partiendo de las características de los distintos tipos de estudios epidemiológicos, buscando una mayor validez y adecuación al objetivo, pero teniendo también en cuenta los recursos humanos, económicos y de tiempo disponibles.

En esta fase hay que contemplar estos puntos:

- Definición de la población de estudio (aspecto crítico).



-
- Determinación del tamaño muestral (cuidado con muestras tan grandes que se escapen a los recursos del investigador).
 - Selección de la muestra (junto con los criterios de inclusión y exclusión, lo que es otro aspecto crítico).
 - Periodo de estudio (no puede ser excesivamente largo).
 - Selección y definición de las variables (otro tema fundamental).
 - Fuentes de información que se van a utilizar (más importante de lo que parece).
 - Procesamiento y análisis de los datos.

Por último, la planificación de las distintas actividades del estudio incluye:

- Estudios pilotos previos para perfeccionar la recogida de datos y probar fallos operativos.
- Entrenamiento de los investigadores en técnicas necesarias (¿medimos todos igual la tensión arterial?).
- Diseñar planes de contingencia
- Formar al personal investigador en el manejo del protocolo de investigación, no asumiendo que todo el mundo lo va a leer.
- Disponer de un cauce organizado para responder dudas.
- Diseñar cuidadosamente el Cuaderno de Recogida de Datos y la base de datos (lo que incluye un soporte electrónico adecuado).
- Entrenar a las personas que teclean datos.
- Diseñar controles de calidad intermedios.

A menudo buenas ideas de investigación se malogran por no cuidar aspectos básicos tan simples, pero a menudo olvidados. Sin duda todo ello redundará en mejorar la calidad de los datos.



No todos los estudios requieren tantas medidas como las descritas, pero según gane complejidad el protocolo de investigación estos aspectos son cada vez más relevantes.

Como se aprende del terreno de la seguridad no hay que dar nada por seguro y es preferible comprobar, repetir y volver a comprobar.

2.2. Criterios de clasificación de los estudios.

Existen varias formas de clasificar los estudios, que no son excluyentes entre sí:

2.2.1. Por su finalidad:

- Estudios descriptivos: describen frecuencia y distribución de aspectos relacionados con la salud. Corresponden a la parte de “medir” del axioma “medir y comparar” del que antes hablamos.
- Estudios comparativos o analíticos: buscan asociaciones causales o riesgos, para lo que debe analizar o comparar al menos dos grupos entre sí. Todo estudio comparativo tiene una fase descriptiva.

2.2.2. Por la unidad de observación:

- Personas: estudios centrados en pacientes o personas (sujetos de investigación, empleando la terminología legal).
- Poblaciones completas: son los denominados estudios ecológicos. Analizan que ocurre en un grupo concreto (en un aula escolar la prevalencia de obesidad es el 40%, y un 80 % tienen vida sedentaria, pero sin asociar persona a persona ambas variables).
- Otros: experimentación animal, estudios con células o tejidos....



2.2.3. Direccionalidad: Este aspecto, muy importante, considera la relación entre causa y efecto, siempre y cuando sea aplicable (no en todos los estudios se puede hacer):

- Desde la causa al efecto: estudios de cohortes. Estos estudios primero miden las causas o factores de riesgo posibles de una enfermedad y luego analizan si ocurre o no la enfermedad a lo largo del tiempo.
- Desde el efecto a la causa: estudios de casos y controles. Miden a la vez el posible factor de riesgo en grupos de enfermos y no enfermos.
- Sin direccionalidad: estudios transversales. En estos casos no tiene sentido aplicar esta terminología.
-

A los primeros (cohortes) les denominamos estudios longitudinales, para diferenciarlos de los transversales, por que miden a lo largo del tiempo como ocurre un hecho que es la enfermedad.

2.2.4. Relación temporal: mide la relación en el tiempo entre el investigador y su investigación en comparación con el hecho que desea medir. De esta forma, los estudios se pueden clasificar como:

- Estudios prospectivos o concurrentes: que considera casos ocurridos o detectados desde el inicio del estudio. Es decir, coinciden en el tiempo investigación y nuevos casos.
- Estudios retrospectivos: considera casos ocurridos o detectados antes del inicio del estudio y que deben ser recopilados de un registro. En estos casos, la calidad del registro es crítica para asegurar la validez del estudio.

2.2.5. Asignación del factor de estudio: Este aspecto es el más importante. Se refiere a aquellos estudios en los que estamos investigando como un factor (un medicamento, una terapia, un factor de riesgo) modifica la enfermedad.



-
- Estudios experimentales: si el investigador asigna mediante un criterio fijado por él la exposición a un factor que modifica la enfermedad (ya sea fármaco o factor de riesgo/protección).
 - Estudios observacionales: el investigador no controla la asignación del factor de estudio. En esos casos es la naturaleza o el comportamiento del sujeto la que asigna el denominado factor de estudio.

En términos de calidad investigadora (si podemos emplear este término), sin lugar a dudas son los estudios experimentales los que más calidad aportan. Por las limitaciones metodológicas que padecen los estudios en ciencias de la salud estos estudios se deberían llamar cuasi-experimentales por que no son exactamente iguales a los estudios experimentales de otros campos de la ciencia, pero es una terminología muy habitual.

Después de los estudios experimentales con pacientes (ensayos clínicos) los siguientes estudios son los longitudinales, más creíbles en principio que los transversales. Siempre es preferible el diseño prospectivo al retrospectivo. Por último, un diseño comparativo siempre ofrece más evidencia que otros descriptivo o no comparativo.



TEMA 3. Tipos de estudios de investigación en Ciencias de la Salud.

3.1 Estudios descriptivos: Son los que tienen por objeto observar y describir la realidad en función de las variables de tiempo, lugar y persona. Son con diferencia los estudios más frecuentes y la aplicación más habitual de la epidemiología.

Algunos estudios descriptivos tienen como unidad de análisis el individuo, entre los que están la descripción de un caso (“a propósito de un caso”), la serie de casos y los estudios transversales. Otros tienen como unidad de análisis la población o el grupo de individuos y son los llamados estudios ecológicos.

- **Serie de casos o “a propósito de un caso”.** Describen uno o varios pacientes con diagnóstico similar, que presentan alguna característica inusual. El paso fundamental en ellos es establecer la definición de caso y los criterios de inclusión y exclusión. Su principal limitación es la falta de un grupo control, lo que impide evaluar hipótesis de tipo causal.

Estos estudios suponen una tercera parte de las publicaciones biomédicas.

En muchos casos quedan en la bibliografía como la descripción de una rareza, pero en algunas ocasiones han sido el primer paso para la identificación de una enfermedad emergente o para un nuevo avance. Un ejemplo de ello fue la descripción en 1981 en Los Ángeles de 5 casos de neumonía por *Pneumocystis carinii* en hombres homosexuales previamente sanos. Esta serie de casos fue la primera descripción del sida y el primer paso para el descubrimiento del virus de la inmunodeficiencia humana.



La descripción de un caso o serie de casos puede plantearse de forma transversal o longitudinal.

- **Estudios descriptivos con datos agregados o ecológicos.**

También son llamados *estudios ecológicos descriptivos*.

Describen la frecuencia de un evento relacionado con la salud en la población, su tendencia en el tiempo, su distribución en el espacio o en función de otras características de las personas como el sexo, edad, profesión, nivel educativo, clase social, etc.

Son abordajes muy frecuentes en el campo de la salud pública, como apoyo a actividades de planificación y gestión sanitarias.

Las posibilidades de estos estudios están condicionadas a la existencia de sistemas de información y a la calidad de sus datos.

Permiten obtener medidas de frecuencia absolutas y relativas, de incidencia, prevalencia o mortalidad, pudiendo establecer comparaciones entre grupos.

Los cambios en la frecuencia del evento estudiado a lo largo del tiempo, las diferencias geográficas o las diferencias en función de las características de las personas, son fuentes de hipótesis para estudios analíticos posteriores.

El planteamiento general de estos estudios es sencillo y requiere pocos medios y tiempo para su ejecución.



Los estudios descriptivos en función de la variable tiempo han encontrado un amplio desarrollo a través de los *análisis de series temporales* y de *edad-periodo-cohorte*.

En este apartado también se incluyen los estudios que exploran posibles agregaciones temporales y/o espaciales de casos.

- **Estudios descriptivos de prevalencia.**

Las encuestas de prevalencia, también llamadas estudios transversales, pueden realizarse con una finalidad analítica, pero habitualmente son descriptivas.

No tienen direccionalidad, es decir, toda la información recogida hace referencia a un mismo momento en el tiempo.

Si la selección de la muestra se hace de forma representativa de la población diana, los resultados serán generalizables a dicha población.

Ofrecen una visión instantánea de lo que ocurre en la población en un momento dado. La medida de frecuencia que proporcionan es la prevalencia.

Entre los estudios descriptivos transversales se incluyen los que evalúan métodos diagnósticos, comparando los resultados de una técnica con los de la técnica de referencia, y también, los estudios que evalúan la concordancia entre técnicas o entre observadores.

3.2. Estudios observacionales de cribado de hipótesis.

Exploran posibles relaciones de causa-efecto, pero por las limitaciones de su diseño sus resultados no son concluyentes e incluso pueden inducir resultados



contradictorios u opuestos a los reales por sesgos desconocidos y por desgracia incontrolables con el diseño.

Sin embargo, por su simplicidad a veces han aportado sugerentes ideas a un coste mínimo.

Incluyen los estudios ecológicos y los de morbilidad o mortalidad proporcional.

Muchos manuales de epidemiología incluyen este grupo de estudios en el apartado de los descriptivos. Sin embargo, Rothman y Greenland se refieren a éstos como estudios de cribado de hipótesis, ya que cumplen la función de preseleccionar hipótesis de trabajo para posteriores estudios analíticos.

-Estudios ecológicos de correlación.

Se caracterizan porque las unidades de análisis son grupos de individuos o poblaciones.

Utilizan una medida de la frecuencia de la enfermedad en cada una de las poblaciones de estudio y evalúan su relación con algún factor de interés como la edad, la fecha de estudio, la utilización de servicios de salud, el consumo de determinados alimentos, de medicamentos o de otros productos.

P. ej: análisis de la correlación a nivel de países entre el consumo per cápita de cigarrillos y la tasa de mortalidad por enfermedad coronaria.

La medida de asociación que se obtiene es el coeficiente de correlación y por ello también se les llama estudios de correlación.

Los grupos de individuos que se utilizan como unidad de estudio son habitualmente los habitantes de distintos países, regiones o ciudades, pero



también pueden compararse cohortes de nacimiento o grupos más pequeños como colegios, empresas o familias. Cuando se compara la población de un mismo lugar en distintos momentos del tiempo se habla de *estudios de tendencia ecológica*.

La principal ventaja de estos diseños es que se realizan sobre fuentes de información disponibles, lo que hace que sean rápidos de ejecución y económicos.

Principal limitación: la imposibilidad de establecer la relación entre la exposición y el efecto a nivel de los individuos. Puede ocurrir que se observe una asociación cuando se analizan poblaciones, sin que exista realmente en el nivel de los individuos, situación que se denomina *falacia ecológica*.

La segunda limitación importante es la dificultad para controlar el efecto de potenciales factores de confusión.

- **Estudios de morbi/mortalidad proporcional.**

Evalúan la asociación entre un factor y la morbilidad o la mortalidad por una determinada causa.

Se clasifican en:

- *Estudios de mortalidad proporcional*, si el efecto es la mortalidad.
- *Estudios de morbilidad proporcional*, si el efecto es la incidencia de una enfermedad.

Se realizan a partir de información de registros.



En un estudio de mortalidad proporcional la comparación se establece entre la proporción de las muertes en sujetos expuestos que es debida a una determinada causa, con la misma proporción en los sujetos no expuestos.

Para algunos autores los estudios proporcionales son un tipo particular de estudios de casos y controles (Rothman y Greenland, 1998), donde los controles son enfermos o fallecidos por otras causas.

La aplicabilidad de estos estudios está limitada en función de las variables que se recogen en los registros. La medida de efecto que se obtiene es la razón de mortalidad (o morbilidad) proporcional.

Esta medida es un buen estimador de la razón de tasas cuando se cumple la asunción de que la morbilidad o mortalidad por las restantes causas es independiente de la exposición; no obstante, esta asunción es difícil de demostrar.

Otra limitación de los estudios de mortalidad proporcional es que no permiten diferenciar si la exposición se asocia a la enfermedad o a la evolución fatal en los que la padecen.

3.3. Estudios analíticos observacionales.

Analíticos son aquellos estudios que pueden evaluar hipótesis sobre las causas de los fenómenos de salud y enfermedad.

Se clasifican en dos grandes grupos: estudios observacionales o experimentales.



En los observacionales el investigador se limita a comparar sujetos en función de su exposición al factor objeto de estudio, sin haber intervenido en dicha exposición.

La mayor parte de los autores distinguen tres tipos de *diseños observacionales básicos*: los de corte transversal, los de casos y controles, y los de cohortes.

Además hay diseños híbridos que combinan elementos de distintos diseños básicos.

- **Estudios de corte transversal.**

En ocasiones evalúan hipótesis de tipo causal.

La muestra suele ser representativa de la población de estudio, lo que facilita que los resultados sean generalizables a la población.

No tienen direccionalidad: tanto la evaluación de la exposición como la del efecto hacen referencia a un mismo momento en el tiempo. Esto impide establecer la relación temporal entre ambos, limitando su utilidad en la investigación de relaciones causales.

Otros problemas son que ante una asociación puede no ser fácil distinguir entre un factor de riesgo y un factor pronóstico, y que los casos con mayor duración tienden a estar sobrerrepresentados.

Por todos estos motivos se les da menor valor que a los otros diseños analíticos a la hora de evaluar asociaciones causales.



No obstante, los estudios transversales pueden ser adecuados para evaluar exposiciones que no cambian en el tiempo, como el tipo sanguíneo, o cuando la exposición actual es un buen indicador de la exposición relevante respecto al posible efecto, lo cual es más probable si la exposición es constante en el tiempo.

- Estudios de casos y controles.

En los estudios de casos y controles se seleccionan dos grupos de sujetos, unos con el efecto que se estudia (casos) y otros sin él (controles).

La direccionalidad del estudio se establece hacia atrás, ya que en un segundo paso se indaga en cuáles de los sujetos estuvo presente la exposición que se estudia.

En función de la relación temporal puede ser:

- *retrospectivo o histórico*, si se incluyen casos prevalentes,
- *prospectivo o concurrente*, si se incluyen sólo casos incidentes, y
- *mixto*, si se incluyen ambos tipos de casos.

Los casos prevalentes están más expuestos a posibles sesgos de supervivencia selectiva y de memoria. En los estudios de casos y controles no es posible obtener medidas de incidencia y se recurre a la *odds ratio* como medida de asociación que estima la razón de tasas de incidencia.

Los estudios de casos y controles son especialmente eficientes para el estudio de enfermedades poco frecuentes o con largos periodos de latencia y cuando es particularmente urgente la generación de hipótesis causales en la investigación de brotes.

- Estudios de cohortes.



La direccionalidad va desde la exposición hasta el efecto.

Los acontecimientos se estudian en la misma secuencia en la que se produjeron.

La selección de la muestra puede hacerse de forma representativa de la población diana o en función de la exposición, como por ejemplo en las cohortes ocupacionales.

Esta última posibilidad es especialmente útil cuando se estudian exposiciones raras. En función de la relación temporal entre los hechos que se estudian y el momento de realización de la investigación, los estudios de cohortes pueden ser históricos, concurrentes o mixtos.

- Los *estudios de cohortes históricas*, también son llamadas *retrospectivas*, se basan en información recogida en registros.
- Los *estudios de cohortes concurrentes o prospectivos* sólo tienen en cuenta los eventos que se producen a partir del comienzo del estudio.
- Los *estudios de cohortes mixtas* son los que combinan información prospectiva y retrospectiva.

Inicialmente todos los sujetos están libres de enfermedad y se diferencian según la exposición. A partir de ese momento se requiere contemplar el seguimiento de todos los sujetos durante un intervalo de tiempo suficiente para dar margen al periodo de inducción del efecto que se analiza.

Al final del seguimiento se evalúa la aparición del efecto en los dos grupos (grupo de expuestos y grupo de no expuestos).



Los estudios de cohortes permiten establecer la relación temporal entre la exposición y el efecto, y obtener medidas de incidencia. Además si las cohortes son concurrentes se minimizan los sesgos en la medida de la exposición.

- **Híbridos.**

Bajo este epígrafe se incluyen los *estudios de cohorte-casos* y los *estudios de casos y controles anidados en una cohorte*.

En ambos se parte de un estudio de cohortes. En el análisis se incluyen todos los casos y, como grupo de comparación, se toma una submuestra de la cohorte. De esta forma se reduce considerablemente el número de sujetos que es necesario estudiar, sin que ello suponga una disminución importante de la potencia del estudio. Estos diseños consiguen mejorar mucho la eficiencia cuando la medición de la exposición o de otras variables de análisis es costosa.

3.4. Estudios experimentales.

Son aquellos en los que la exposición es asignada por el investigador en función de lo establecido en el protocolo.

Con el fin de garantizar la mayor comparabilidad entre los grupos y evitar sesgos, lo ideal es que la asignación de la exposición se realice de forma aleatorizada, aunque esto no siempre es posible.

Muchos autores consideran estudios experimentales exclusivamente a aquellos realizados de forma aleatorizada: ensayos clínicos, ensayos de campo y ensayos comunitarios.



Por razones éticas estos estudios únicamente tienen aplicación para evaluar el efecto de exposiciones potencialmente terapéuticas o preventivas. Además, las alternativas que se comparan deben ser a priori igualmente aceptables en función del conocimiento que se tiene sobre el tema; por ejemplo, el mejor tratamiento existente debe ser la referencia de comparación para cualquier nuevo tratamiento.

- ***Estudios cuasi-experimentales o de intervención no aleatorizados***

Son aquellos en los que hay asignación controlada de la exposición sin seguir un procedimiento aleatorio.

Las unidades de asignación del factor de estudio pueden ser los individuos o también las poblaciones.

Dentro de estos estudios se incluyen los estudios en los que la comparación se establece entre los mismos sujetos antes y después de aplicar la intervención.

Se utilizan para la evaluación de la eficacia de actividades preventivas y de educación para la salud.

- ***Ensayo clínico***

También llamados ensayos aleatorizados controlados, las unidades de estudio son pacientes.

Su objetivo es evaluar nuevos tratamientos o estrategias de prevención secundaria. Sobre la población diana se aplican los criterios de inclusión y exclusión, previamente protocolizados, y se obtienen los sujetos elegibles. Existe



un acuerdo universal de que la asignación aleatoria del tratamiento es el mejor método para garantizar la comparabilidad inicial de los grupos respecto a todas las características.

Cuando es posible, conviene emplear enmascaramiento, es decir, que el paciente, la persona que asigna la exposición y la que evalúa el resultado, desconozcan qué pacientes reciben el tratamiento nuevo y cuáles el tratamiento habitual (o el placebo).

Finalizado el seguimiento de los pacientes se procede a comparar los resultados de la intervención entre los grupos.

- ***Ensayo de campo***

Evalúan medidas de prevención primaria y, por tanto, se aplican en sujetos sanos.

Como la incidencia de la mayoría de enfermedades en población sana es muy baja, se requiere que estos estudios tengan un elevado número de participantes, para poder obtener resultados concluyentes.

Esto limita la aplicabilidad de los ensayos de campo a medidas de prevención en enfermedades frecuentes o muy graves y hace que sean estudios económicamente muy costosos.

P. ejemplo: la evaluación del suplemento vitamínico en población sana para prevenir ciertas enfermedades.

- ***Ensayo de intervención comunitaria***



Los ensayos de intervención comunitaria evalúan también medidas de prevención primaria, pero se caracterizan porque la unidad de asignación del factor de estudio y de análisis son grupos de individuos o comunidades.

Algunos ejemplos de posibles intervenciones comunitarias son la fluoración de las aguas de abastecimiento público, los programas de educación utilizando medios de comunicación de masas, o la intervención medioambiental en el lugar de trabajo.

La asignación puede hacerse de forma aleatoria entre los grupos o comunidades, pero lo más frecuente es que no sea así y en este caso pasarían a ser estudios cuasi-experimentales.

3.5. Grado de evidencia científica aportado por los estudios:

A la hora de seleccionar un diseño, la evidencia es inversamente proporcional a la simplicidad. Al grupo de investigación que se estrene se le debe recomendar empezar por los más sencillos, aunque la evidencia aportada sea mínima.

Como puede desprenderse de las guías de terapia o de las recomendaciones la evidencia proporcionada por algunos de estos estudios es despreciada por la comunidad científica.

Podemos decir, que por orden decreciente, la “calidad de la evidencia” aportada por los estudios es:

1. Ensayo controlado aleatorizado con enmascaramiento.
2. Ensayo controlado aleatorizado sin enmascaramiento.
3. Ensayo controlado no aleatorizado, estudio de intervención
4. Estudio de cohortes concurrentes (prospectivas)



5. Estudio de cohortes históricas (retrospectivas)
6. Estudio de casos y controles con casos incidentes
7. Estudio de casos y controles con casos prevalentes
8. Estudio de corte transversal
9. Estudio de mortalidad o morbilidad proporcional
10. Estudio ecológico de correlación

TEMA 4. ESTUDIOS SOBRE PRUEBAS DIAGNÓSTICAS.

1.1. Sensibilidad, especificidad y valores predictivos.

La validez es la capacidad de la prueba de discriminar entre los casos y los no casos para una enfermedad o condición clínica.

La validación de una prueba diagnóstica se realiza mediante la comparación de sus resultados con los obtenidos con el mejor instrumento de medida o con el verdadero resultado si es conocido.

	Verdaderos enfermos	Verdaderos Sanos	
Diagnosticados por la prueba	A (verdaderos +)	B (falsos +)	A+B
No diagnosticados por la prueba	C (falsos negativos)	D (verdaderos negativos)	C+D

La *sensibilidad* de una prueba es la probabilidad de que un sujeto con la enfermedad sea adecuadamente clasificado.



$$\text{Sensibilidad} = A / (A+C)$$

Una prueba será tanto más sensible cuanto menos falsos negativos (FN) proporcione. En las pruebas de cribado generalmente se utilizan test de alta sensibilidad para captar a casi todos los enfermos. También en enfermedades graves tratables, en las que no debe permitirse la pérdida de ningún caso.

La *especificidad* de una prueba es la probabilidad de que un sujeto sano sea clasificado correctamente. Una prueba será tanto más específica cuanto menos falsos positivos (FP) proporcione.

$$\text{Especificidad} = D / (B+D)$$

En general, las pruebas de confirmación diagnóstica deben tener una elevada especificidad, así como en la identificación de enfermedades graves no tratables, en las que un resultado falso positivo puede producir serios trastornos al individuo.

El *valor predictivo positivo* de una prueba es la probabilidad de que un sujeto con un resultado positivo sea un individuo enfermo.

$$\text{VPP} = A / (A+B)$$

El *valor predictivo negativo* es la probabilidad de que un sujeto con un resultado negativo en la prueba sea un individuo sano.

$$\text{VPN} = D / (C+ D)$$



La *seguridad* de una prueba diagnóstica se valora mediante el valor predictivo positivo y negativo, ya que nos dice en qué grado los resultados positivos y negativos de la prueba son correctos.

1.2. La influencia de la prevalencia de la enfermedad en la población en los resultados de una prueba diagnóstica empleada para su identificación.

Una prueba diagnóstica de calidad debe tener una sensibilidad y especificidad altas (próximas al 100%), pero también deben tenerse en cuenta las condiciones de aplicación de la prueba.

La prevalencia de enfermedad en la población en la que se aplica el test, aunque no modifica la sensibilidad y especificidad puede modificar los valores predictivos positivos y negativos convirtiendo a una prueba de calidad en una intervención diagnóstica inútil.

Si se aplica una prueba con una alta sensibilidad y especificidad, como el test de ELISA para la detección del virus del SIDA, en una población con una prevalencia de seropositividad al VIH muy baja se examinará a muchos sujetos sin la enfermedad dando lugar a varios falsos positivos.

Veamos un ejemplo de lo mismo con datos con alta prevalencia de la enfermedad estudiada:

	Verdaderos Enfermos	Verdaderos sanos	
Prueba +	90	100	190
Prueba -	10	900	910

Sensibilidad = $90 / 100 = 0,90$



$$\text{Especificidad} = 900 / 1000 = 0,90$$

$$\text{VPP} = 90 / 190 = 0,47$$

$$\text{VPN} = 900 / 910 = 0,99$$

$$\text{Prevalencia real de la enfermedad} = (90+100) / (10+900) = 0,21$$

Y ahora otro ejemplo con datos de menor prevalencia:

	Verdaderos Enfermos	Verdaderos sanos	
Prueba +	90	1000	1090
Prueba -	10	9000	9010

$$\text{Sensibilidad} = 90 / 100 = 0,90$$

$$\text{Especificidad} = 9000 / 10000 = 0,90$$

$$\text{VPP} = 90 / 1090 = 0,08$$

$$\text{VPN} = 9000 / 9010 = 0,99$$

$$\text{Prevalencia real de la enfermedad} = (90+1000) / (10+9000) = 0,12$$

Como puede verse sensibilidad y especificidad no cambian pero los valores predictivos sufren cambios dramáticos.

Una de las consecuencias que tiene este hecho es que cuando se emplea una prueba diagnóstica en una población de alta prevalencia el resultado es altamente fiable, pero cuando se emplea un test en poblaciones de baja prevalencia (por ejemplo, para hacer un screening de identificación de casos precoces de cáncer de mama en una población sana) hay que asumir un coste de falsos positivos que puede ser prohibitivo y hace poco ético el uso de la prueba y nada rentable en términos económicos por que obliga a descartar una gran cantidad de falsos positivos con nuevas pruebas.



1.3. Las razones de máxima verosimilitud (RMV).

Un indicador útil para discriminar entre pruebas diagnósticas por su valor informativo es la llamada en los libros escritos en inglés “likelihood ratio” que puede traducirse como razón de máxima verosimilitud o también llamada razón de probabilidades.

Este indicador no tiene un umbral que establezca cuando la prueba diagnóstica tiene poco valor, pero puede ser útil para conocer los resultados que aportará la realización de la prueba.

La razón de máxima verosimilitud del resultado positivo de una prueba diagnóstica se calcula mediante la siguiente fórmula:

$RMV (+) = \text{sensibilidad} / (1 - \text{especificidad})$.

Este cociente indica las probabilidades de dar + en la prueba si se es un verdadero enfermo en comparación con las probabilidades de dar + si se ha descartado la enfermedad.

Cuanto más elevado es el resultado de la fórmula anterior mejor es la prueba. Por ejemplo, se ha estimado que una tomografía computarizada abdominal para el diagnóstico de una lesión esplénica tiene una sensibilidad del 95% y una especificidad del 96%, resultando una razón de máxima verosimilitud de un resultado positivo de 23,8. Sin embargo, los termogramas para la detección del cáncer de mama en la mujer tienen una sensibilidad del 61% y una especificidad del 74%, dando una razón de máxima verosimilitud de un resultado positivo baja (2,35).



La representación gráfica del numerador y denominador de la razón de máxima verosimilitud puede servir de ayuda para discriminar cuál es el punto que ofrece mejores resultados. El siguiente apartado se dedica a la curva ROC, gráfico estándar que representa los puntos de equilibrio entre sensibilidad y especificidad.

La razón de máxima verosimilitud del resultado negativo de una prueba se calcula mediante la siguiente fórmula

$$\text{RMV}(-) = \text{especificidad} / (1 - \text{sensibilidad}).$$

En este caso, cuanto más bajo es el valor, mejor es la prueba.

1.4. El umbral de positividad de una prueba. La curva de características operativas (curva ROC).

La aplicación de una prueba diagnóstica implica tomar una serie de decisiones que tendrán consecuencias en su evaluación.

Un test no aporta la misma información diagnóstica si el umbral para considerar la prueba positiva es bajo que si es muy exigente. La glucemia como prueba diagnóstica de diabetes presenta distintos valores de sensibilidad y especificidad según el umbral diagnóstico sea 100 mg/dl o 140 mg/dl.

Si se considera este último umbral, ciertamente casi todas las personas con una prueba positiva (>140 mg/dl.) estarán enfermas.

Sin embargo, muchas otras personas con diabetes no serán calificadas como enfermas con una prueba tan rigurosa. La prueba sería entonces muy específica.



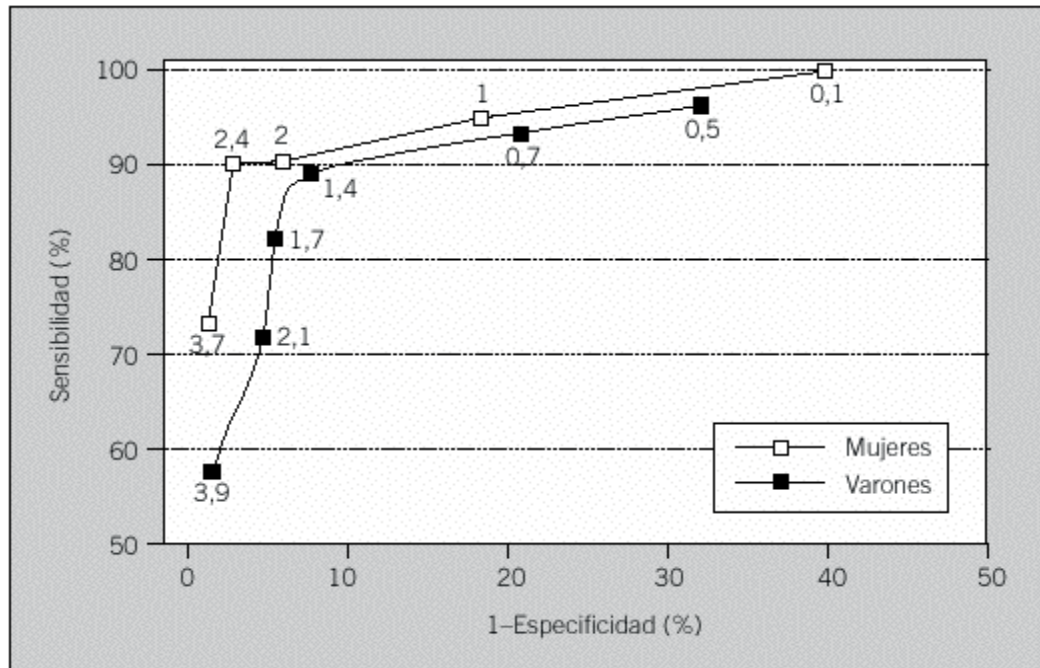
Situándose en el otro extremo, y considerando que todos los pacientes con una glucemia mayor de 70 mg/dl. fueran diagnosticados como diabéticos, muy pocas personas con la enfermedad dejarían de ser diagnosticadas, pero la mayoría de la gente normal sería etiquetada como diabética. La prueba sería entonces muy sensible pero poco específica.

Este balance entre sensibilidad y especificidad debe considerarse a la hora de definir el umbral adecuado para considerar positiva la prueba.

La construcción de una curva ROC (Receiver Operator Characteristic Curve) es una herramienta muy útil en esa decisión.

Esta curva se construye con los diferentes resultados que se obtienen en una prueba diagnóstica al colocar el umbral de positividad en diferentes puntos.

En la curva se representan en un gráfico los pares sensibilidad y 1-especificidad, que en definitiva son las razones de máxima verosimilitud de un resultado positivo. El eje de abscisas crece conforme disminuye la especificidad y el eje de ordenadas crece conforme aumenta la sensibilidad.



En la [figura](#) se muestra un ejemplo de curva ROC para una prueba diagnóstica de nefropatía diabética, el cociente albúmina/creatinina en orina.

El punto de corte o umbral de positividad que más se aproxima al extremo superior izquierdo es el que más área deja debajo de la curva, y corresponde con el de mayor razón de máxima verosimilitud. En el caso de las mujeres el mejor punto de corte es 2,4 g/mol, y corresponde con una sensibilidad de 90% y una especificidad de 97%. En el caso de los hombres es 1,4 g/mol (sensibilidad 89% y especificidad 92%).

Además de poder realizar un análisis cualitativo visualmente, las curvas ROC permiten un análisis cuantitativo preciso para la evaluación de la exactitud de una prueba diagnóstica. El punto que deja mayor área debajo de la curva es el de mayor razón de máxima verosimilitud, y por lo tanto en el que se ofrece una mayor exactitud global.

Visualmente se puede discriminar cuál es el punto que se aproxima más al extremo superior izquierdo, pero para mayor exactitud se calcula el área debajo



de la curva. Ésta se interpreta como la probabilidad de clasificar correctamente un par de individuos sano y enfermo, seleccionados al azar de la población objeto de estudio, al aplicarles la prueba diagnóstica en cuestión. Es decir, la probabilidad de que el resultado de la prueba resulte más anormal en el sujeto enfermo que en el sano.

El área debajo de la curva es una medida global de la exactitud que permite la comparación estadística entre pruebas diagnósticas. Los valores están comprendidos entre 0,5, el peor resultado posible y 1, el mejor. Un área debajo de la curva de 0,72 significa que un individuo seleccionado aleatoriamente del grupo de enfermos obtendrá el 72% de las veces un valor mayor en la prueba diagnóstica que un individuo elegido al azar del grupo no enfermo.

2. FIABILIDAD DE UNA PRUEBA DIAGNÓSTICA.

2.1. Conceptos generales.

Hay algo de variabilidad en cualquier procedimiento de medición, particularmente cuando hay una valoración afectada por la subjetividad de un observador.

La variabilidad no puede ser eliminada totalmente, pero el conocer sus causas y su evaluación es necesario para garantizar su calidad.

El grado de variabilidad puede ser evaluado determinando su validez o exactitud, aspecto abordado anteriormente, y además mediante la evaluación de la fiabilidad.

La *fiabilidad o precisión* es la capacidad de una prueba diagnóstica de obtener resultados comparables o replicables si se utiliza en condiciones idénticas.



Las fuentes de variación se pueden resumir de varias maneras, pero vamos a adoptar una clasificación relativamente habitual en la literatura española.

Las fuentes u orígenes de la variabilidad poder ser originadas por:

— La *variabilidad intraobservador*: el mismo profesional sanitario puede obtener diferentes resultados al repetir una medición en el mismo paciente.

— La *variabilidad interobservador*: las mediciones realizadas por diferentes profesionales suelen ser diferentes, lo cual puede estar motivado por la formación del profesional, la experiencia, la motivación, la agudeza visual o auditiva y otras características personales que pueden afectar en los resultados de las observaciones.

— La *variabilidad biológica* inherente a la característica que se desea medir. Por ejemplo, la tensión arterial está sujeta a cambios en dependencia de la situación en que se encuentre el paciente en el momento de la toma (ingesta previa, estrés, condición física, etc.).

— La *variabilidad del instrumento de medida*. Puede estar motivado por la calibración, las unidades en que se mide o la influencia sobre el propio instrumento de factores ambientales como el nivel de ruido o la iluminación.

— La *variabilidad aleatoria*. Es el grado de variación que no puede atribuirse a los factores anteriores. La repetición de una prueba suele aumentar la fiabilidad al disminuir el error aleatorio.

Otro fenómeno reseñable es el denominado “regresión a la media” que es el efecto de disminución de los valores extremos que se produce al realizar repetidamente una prueba, de tal forma que los valores se van aproximando



cada vez más a su valor medio real traduciendo un fenómeno biológico que podríamos asimilar a la homeostasis, por el cual los organismos biológicos tienden a buscar una situación de equilibrio regulando los valores extremos.

Este hecho es muy importante a la hora de interpretar los resultados de ensayos clínicos pues a veces el efecto supuesto de la medicación no es más que la regresión a la media.

Sin duda, la variabilidad es uno de los grandes enemigos del investigador, por eso su reducción es una de las prioridades del protocolo.

La reducción de la variabilidad hasta valores aceptables debe hacerse actuando sobre las diferentes fuentes de variación identificadas. La estandarización del procedimiento de medición puede contribuir a reducir la variabilidad homogeneizando las condiciones de aplicación del procedimiento diagnóstico o entrenando de forma similar a los distintos observadores.

Otra forma de reducir la variabilidad es repitiendo la medición en varias ocasiones para tomar la media de las mismas como valor verdadero del sujeto estudiado. Si la variabilidad es extrema no es descabellado realizar hasta 20 mediciones empleando la mediana de los valores obtenidos y no la media.

La evaluación de la fiabilidad de una prueba diagnóstica se realiza valorando cada una de las fuentes de variabilidad.

La variabilidad interobservador: se mide la concordancia entre los resultados obtenidos por los distintos profesionales, manteniendo constantes el resto de los factores que pueden influir en los resultados.



En dependencia del carácter de la variable que se mide se utilizará un método de evaluación de la variabilidad o concordancia diferente.

Cuando se trate de variables nominales o binarias (enfermo/sano) se utilizará el coeficiente Kappa. Si las variables son ordinales (leve/moderado/ grave) se utilizará el coeficiente Kappa ponderado. Si las variables sean cuantitativas se utilizará el coeficiente de correlación intraclass.

2.2. Coeficiente kappa.

Ejemplo: una prueba diagnóstica de detección de sangre oculta en heces. Los resultados de la prueba pueden ser negativos (-) o positivos (+). La prueba fue realizada por dos médicos (A y B) en 110 pacientes.

	Resultado real (-)	Resultado real (+)	Total
Medico A	48	62	110
Medico B	47	63	110
	95	125	220

A la vista de la tabla anterior, los resultados parecen bastante concordantes, pero ¿coinciden los diagnósticos de ambos médicos? Veamos esta nueva tabla:

	Medico A (-)	Medico A (+)	total
Medico B (-)	37	10	47
Medico B (+)	11	52	63
	48	62	110



En esta tabla podemos definir:

Índice de Acuerdo Global: $(37+52) / 110 = 0,81$ (proporción de completas concordancias entre ambos médicos).

Aunque parece alta, debemos considerar que al ser el resultado dicotómico, la concordancia al azar puede ser muy elevada. El índice kappa calcula la concordancia NO debida al azar.

Calculemos el Índice de Acuerdo Atribuible al Azar (IAAA), empleando una técnica similar a la del cálculo de la Chi2 para el cálculo de los esperados

	Medico A (-)	Medico A (+)	total
Medico B (-)	20,51	26,47	47
Medico B (+)	27,49	35,51	63
	48	62	110

El valor de cada celda se obtiene de multiplicar la suma total de su fila y columna correspondiente y dividiendo por n (110).

Ejemplo: Para 20,51 = $(48 \times 47) / 110$.

IAAA = $(20,51 + 35,51) / 110 = 0,51$.

Por tanto, el 51 % de las observaciones concordantes se deben al azar.
Este valor es muy alto.

Finalmente para el cálculo del Kappa se empleará este dato:



$$Kappa = (IAS - IAAA) / (1 - IAAA) = (0,81 - 0,51) / (1 - 0,51) = 0,30 / 0,49 = 0,61.$$

Solo un 61 % del acuerdo no se debe al azar.

El Kappa oscila de -1 (máximo desacuerdo) a + 1 (máximo acuerdo). Se considera bajo o muy bajo (por tanto no aceptable) valores por debajo de 0,4, mediano de 0,4 a 0,75 y bueno por encima de 0,75.

El coeficiente *Kappa Ponderado* se utiliza cuando las variables son ordinales, de tal forma que haya más de dos categorías de valoración con un orden jerárquico entre ellas. Se calcula de forma análoga al coeficiente Kappa Simple, pero multiplicando el valor de cada casilla por un peso que pondere la magnitud del desacuerdo, ya que no es lo mismo discrepar entre leve y moderado, que entre leve y grave.

2.3. El Coeficiente de Correlación Intraclases.

El Coeficiente de Correlación Intraclases (CCI) se utiliza para valorar la coincidencia cuando las variables medidas son cuantitativas. Es incorrecto utilizar el Coeficiente de Correlación de Pearson.

Ejemplo: dos médicos toman la tensión arterial sistólica de forma consecutiva y con el mismo esfigmomanómetro a 10 pacientes.

Medición de la tensión arterial sistólica.

Paciente	Médico A	Médico B	Diferencia (B-A)
1	135	140	5
2	140	145	5
3	130	135	5
4	145	150	5



5	140	145	5
6	150	160	10
7	140	145	5
8	135	140	5
9	140	145	5
10	135	145	10
Media	139	145	6
Des. Estándar	5,68	6,67	2,11

Como puede apreciarse en la [tabla](#) las mediciones de los dos médicos son diferentes y los valores registrados por el médico B son en todos los casos superiores a los registrados por el médico A, y además la diferencia es de igual magnitud casi siempre (en 8 de las 10 ocasiones la diferencia es de 5 mm de Hg, y en las dos restantes la diferencia es de 10 mm de Hg). Las diferencias son sistemáticas, van siempre en la misma dirección y son de similar magnitud.

En este caso el coeficiente de Pearson detecta una correlación de 0,95. Sin embargo, el CCI sí que detecta que las diferencias son sistemáticas y el resultado refleja una concordancia de 0,64.

El cálculo del CCI se basa en un modelo de análisis de la varianza de medidas repetidas.

Se descompone la variación total observada (varianza) en los siguientes componentes: la variabilidad debida a las diferencias entre pacientes o individuos observados (Varianza entre individuos = σ_E); la varianza intra-individuos (σ_I), en este caso debida a diferencias en mediciones repetidas realizadas por distintos observadores, y por último la varianza residual (σ_R), que es la parte de variabilidad inexplicable y que atribuimos al azar.



El CCI se define como el cociente entre la variabilidad entre-individuos y la variabilidad total

$$CCI = \sigma_E / (\sigma_E + \sigma_I + \sigma_R)$$

Cuando una prueba diagnóstica sea muy fiable o cuando la concordancia entre dos observadores sea muy alta, los resultados de las mediciones repetidas sobre un mismo individuo serán muy parecidas y, en consecuencia, la varianza intra-individuos será muy baja y el

valor del CCI será alto. Cuanto más se aproxime a 0 el valor de la varianza intra-individuos, más se acercará a 1 el valor del CCI y viceversa. Puesto que es una proporción, los valores del CCI pueden variar de 0 a 1, donde 0 indica ausencia de concordancia, y 1 concordancia o fiabilidad absoluta. Como norma general, valores inferiores a 0,4 indican escasa fiabilidad, entre 0,4 y 0,75 la fiabilidad es mediana, y valores superiores a 0,75 reflejan un grado excelente de fiabilidad.

2.4. Gráficos de Bland-Altman.

Cuando se quiere representar gráficamente los valores de una variable medida por métodos diferentes no es correcto emplear un gráfico tipo nube de puntos situando en el eje X un sistema de medición y en el eje Y el otro.

Lo adecuado es representar estos valores como dos epidemiólogos que dan nombre al gráfico propusieron.

Cada sujeto se representa con un punto, en efecto, pero de esta forma:

En el eje de las X se representa la media de los valores de un sujeto medidos con los dos métodos. En el eje de las Y se representa la diferencia entre ambos valores.



Lo correcto es que los puntos representen una banda estrecha con la misma anchura a lo largo del gráfico. Si se ve una forma de cuerno o diábolo es que la concordancia entre los dos sistemas de medición es mejor en unos valores que en otros (por ejemplo, en la medición de la glucemia capilar y la venosa, la concordancia puede ser buena por debajo de 180 mg/dl, pero por encima de ese valor ya no es tan buena).

Los puntos formarán una forma de cuerno más ancho hacia la derecha del gráfico.

