

7



Cuestiones estadísticas básicas.

Domingo Pérez Flores

1. LA INCERTIDUMBRE EN LA TOMA DE DECISIONES EN MEDICINA.

2. LA ESTADÍSTICA EN LA LECTURA CRÍTICA DE LA LITERATURA.

2.1. Muestreo.

3. INTERVALO DE CONFIANZA.

4. SIGNIFICACIÓN ESTADÍSTICA.

5. MEDIDAS DEL EFECTO DEL TRATAMIENTO.

- 5.1. Reducción absoluta del riesgo.
- 5.2. Riesgo relativo.
- 5.3. Reducción relativa del riesgo.
- 5.4. Número necesario de pacientes para tratar.
- 5.5. Número necesario de pacientes para perjudicar.
- 5.6. Odds ratio.

6. BIBLIOGRAFÍA.

7



1 LA INCERTIDUMBRE EN LA TOMA DE DECISIONES EN MEDICINA

En las últimas décadas ha habido un fuerte y creciente interés por la Estadística en el campo médico. La Medicina es un área de la Ciencia donde los principios y métodos estadísticos han tenido un éxito especialmente consistente y continúan teniéndolo, y ello es debido, fundamentalmente, al importante papel que dichos procedimientos juegan en todos sus ámbitos, tanto para una mejor investigación y comprensión de todas las Ciencias Biomédicas como para una práctica eficaz de cualquiera de las profesiones sanitarias. Es más, en algunos aspectos de la Medicina como, por ejemplo, la Epidemiología, la Salud Pública, los Ensayos Clínicos ó la Medicina Basada en la Evidencia, los procedimientos estadísticos no son sólo necesarios, sino centrales. La visión médica del mundo tiene fuertes aspectos estadísticos¹: a preguntas frecuentes en el campo médico del tipo ¿cómo de frecuente es la enfermedad?, ¿quién es especialmente propenso a contraerla?, ¿cuáles son las posibilidades de recuperación ó de supervivencia?, se responde con procedimientos estadísticos.

Por otra parte, los conocimientos de Estadística le proporcionan al médico una buena capacidad de razonamiento y de criterio tanto en la aplicación del conocimiento al análisis e interpretación de los datos y signos clínicos, como en la definición de la naturaleza de los problemas y la planificación y la ejecución de una estrategia para resolverlo, y, por otra parte, una comprensión de la contribución de la metodología de la investigación y una aptitud para interpretar y aplicar la investigación de otros en su labor diaria.

Pero ¿qué fenómeno hace que la Estadística tenga esta importancia en el desarrollo de la Medicina actual?. Sin duda, el hecho de que en la Medicina, y en las Ciencias de la Vida en general, se da un fenómeno característico, como es la presencia intrínseca y continua de la variabilidad en los seres vivos, que hace que, en su ámbito, se están tomando continuamente decisiones en ambiente de incertidumbre y, como consecuencia, que la Medicina sea, en sí misma, de naturaleza probabilística.

La Medicina, como disciplina, trata con individuos que presentan variaciones en sus diferentes características, que constituyen un fenómeno que se conoce como variabilidad. Estas variaciones son de distintos tipos y dependen de factores biológicos como la edad, el sexo, la constitución genética, etc..., de factores ambientales como, por ejemplo, la dieta, el estrés, el esfuerzo, el estilo de vida, etc..., y de la disponibilidad de recursos sanitarios.

Por una parte, todos los seres vivos muestran variación en sus caracteres y medidas. Casi todas las características encontradas en las Ciencias de la Salud, bien sean ambientales, fisiológicas, bioquímicas ó inmunológicas, etc..., presentan variación. Cada ser humano es único y cada reacción de cada ser humano a cada estímulo ambiental ó a un tratamiento es diferente. Los principios y métodos estadísticos nos permiten medir, describir y analizar esta variabilidad.

Por otra parte, la prestación de la asistencia sanitaria se basa bastante, hoy en día, en los datos clínicos y en las pruebas de laboratorio, características del ser humano que en muchos casos presentan una enorme variabilidad y para las que una función importante de la Medicina es definir los límites normales de esa variación. Es fundamental una base sólida en métodos estadísticos para el análisis de los datos clínicos recogidos y para comprender las posibles falacias en la interpretación de los mismos. El médico debe saber medir e interpretar la sensibilidad y especificidad de los tests de laboratorio que utiliza, así como su repetibilidad, y también tiene que saber como descubrir y controlar las fuentes de variación (intra e inter observador). El conocimiento y la consideración de la variabilidad fisiológica y patológica del dato clínico a que se refiere un informe del laboratorio tienen un componente estadístico que es de gran ayuda tanto para el diagnóstico como para la evaluación de la respuesta al tratamiento. Estos parámetros estadísticos observados tanto en los datos clínicos como en las pruebas de laboratorio son, pues, esenciales para la toma de decisiones racionales en la práctica clínica.

Teniendo en cuenta esta variabilidad del ser humano, las decisiones en la prestación de la atención sanitaria se basan en la experiencia previa con otros pacientes ó con otras comunidades de características biológicas y sociales parecidas. Debido a esta variabilidad, los resultados de esas decisiones no se pueden predecir con exactitud, es decir, estos resultados siempre estarán acompañados por la incertidumbre. Por ello, dado que la Estadística es la herramienta fundamental para tratar con la variabilidad y la incertidumbre, su uso es esencial en la toma de decisiones juiciosas, basadas en criterios racionales, para la prestación de la atención sanitaria.

Vemos, pues, que una de las claves de la Medicina es la toma de decisiones en ambiente de incertidumbre, tanto a nivel de pacientes individuales como a nivel de comunidad. En las tres actividades fundamentales del clínico, como son diagnosticar correctamente las afecciones de sus pacientes, evaluar la gravedad y el desenlace probable de la enfermedad e influir en el curso de ésta de la manera más eficaz y humana posible, es decir, en el diagnóstico, el pronóstico y el tratamiento, la incertidumbre no deja de estar presente y como tal, son actividades con componente estadístico.

En el diagnóstico, el médico ha de decidir a qué categoría pertenece la enfermedad de su paciente. Sea ó no sea él consciente de ello, esta decisión se basa en dos procesos: en primer lugar, definir la categoría de diagnóstico, entendiéndose como tal cuando determinada combinación de síntomas y signos aparece conjuntamente con mucha mayor frecuencia de la que permite la simple coincidencia y cuando esa agrupación estadística, repetible y válida, facilita una base para el pronóstico ó el tratamiento y puede interpretarse en el marco general de los conocimientos existente en cada momento, en materia de ciencias médicas. En segundo lugar, la asignación de la enfermedad del paciente a una categoría determinada, lo que implica cierto riesgo de error dado que ese paciente no tendrá, probablemente, todos y cada uno de los síntomas y signos que se describen como típicos de esa categoría de diagnóstico y sí puede presentar algunos síntomas que tradicionalmente no se asocian con la enfermedad cuya existencia se sospecha. Los procedimientos probabilísticos, estadísticos y de inducción lógica permiten medir el grado de certidumbre de un diagnóstico, mejorar cualitativamente la toma de una decisión sobre un diagnóstico en función de los síntomas y optimizar, en algunos casos, el procedimiento diagnóstico relativo a los síntomas observados y a los costes.



En el pronóstico, el médico debe hacer la valoración ó predicción del posible resultado de un programa de intervención en una comunidad ó de un tratamiento en la enfermedad de un paciente, a la vista de los signos, síntomas y circunstancias. La estimación del riesgo que la enfermedad entraña para la vida ó para el estado de salud ulterior no deja de ir acompañada de incertidumbre y sólo se puede hacer basándose en la experiencia de muchos casos de la misma enfermedad.

En cuanto al tratamiento, la elección del programa de intervención comunitaria adecuado ó del tratamiento apropiado para un paciente se basa, ó debería basarse, en procedimientos estadísticos, pues es una decisión en la que subyace, de nuevo, la incertidumbre, ya que un médico raramente sabe con exactitud como responderá un paciente a un tratamiento determinado, pues siempre hay variaciones en las reacciones de pacientes diferentes y fluctuaciones imprevisibles en el proceso mismo de la enfermedad. El médico elige una intervención ó un tratamiento específicos sobre la base de su propia experiencia, de los resultados obtenidos en pacientes ó comunidades similares que han recibido esa intervención y en el examen de publicaciones sobre los distintos tipos de tratamiento en ensayos clínicos controlados.

7

→ 2

LA ESTADÍSTICA EN LA LECTURA CRÍTICA DE LA LITERATURA

En los inicios de la Medicina, las decisiones sobre métodos de diagnóstico y sobre tratamientos estaban basadas en la autoridad, donde la opinión de un eminente profesional, basada en su experiencia y en la autoridad previa de otros profesionales, era la fuente principal de información. Con el paso del tiempo, conforme la Medicina se hacía más científica, la experiencia se fue formalizando cada vez más en forma de experimentos controlados y estudios más planificados, llegándose a la situación actual en la que la fuente principal para la toma de decisiones del médico sobre métodos de diagnóstico ó de tratamiento es el uso de los datos científicos publicados, es decir, estamos en la era de la Medicina basada en la evidencia. Se ha producido, pues, un cambio cualitativo importante en el paradigma de la Medicina: se ha pasado de la práctica médica basada en la autoridad a la práctica basada en la evidencia, donde la revisión sistemática y la interpretación de toda la evidencia publicada son aceptadas como la base del enfoque actual de la atención sanitaria. El mismo proceso está ocurriendo en otros campos de las Ciencias de la Salud, hablándose en la actualidad de Enfermería basada en la evidencia, Fisioterapia basada en la evidencia, etc..

El profesional sanitario puede enfocar su inspección de la evidencia en varios niveles. El nivel básico y fundamental es la revisión de un artículo científico en el que se informa de los resultados de un estudio concreto y que es importante para estudios relacionados con la propia actividad de este profesional. El siguiente nivel es el artículo de revisión, donde están recogidos conjuntamente la evidencia y la opinión sobre un asunto. Un tercer nivel es el meta-análisis, que combina todos los estudios disponibles sobre un tema, dando una estimación conjunta de algún parámetro de interés, como por ejemplo, el efecto de un tratamiento, etc.. El último nivel serían los proyectos de ámbito mundial, como la Cochrane Collaboration, que proporcionan meta-análisis de todos los tratamientos conocidos, bajo todo tipo de condiciones, los actualiza de forma regular y los distribuye a la comunidad científica. En todos estos niveles es necesario que el profesional sanitario inspeccione los informes aportados en las revistas

médicas, para cuya comprensión necesita dos cosas: una es un conocimiento del tema que se está investigando, como la enfermedad que se intenta tratar, los fármacos que se están prescribiendo, las distintas posibilidades de tratamiento, etc... La otra es una familiarización con los métodos de investigación que se están usando, particularmente los referidos al diseño del estudio y a la información estadística recogida en el artículo, en términos de resumen de datos y análisis e interpretación de resultados, lo que muestra la importancia de los métodos estadísticos en la práctica médica basada en la evidencia.

La decisión sobre qué tratamiento aplicar a un paciente se basa, fundamentalmente, en la observación de la eficacia relativa de las diferentes formas de terapia. En el proceso mediante el que se llega a la decisión interviene, de manera esencial, la evaluación estadística de la información numérica recogida en la literatura existente. En este aspecto, en cualquiera de los cuatro niveles referidos anteriormente, el clínico debe ser capaz, no sólo de comprender los resultados publicados sobre las distintas terapias, sino también de evaluar críticamente su validez, sabiendo detectar cualesquiera razones estadísticas que haya para dudar de las conclusiones. Altman^{2,3,4} señala entre estas razones, como más importantes, las siguientes:

- Uso de porcentajes muestrales como poblacionales.
- Uso de muestras demasiado pequeñas.
- Elección de grupos de comparación inadecuados y grupos control que no lo son en el sentido estadístico.
- Mezcla de pacientes no comparables (hombres y mujeres, jóvenes y ancianos, etc.,).
- Naturaleza sesgada de la muestra de pacientes que recibe el tratamiento.
- Violación de condiciones del análisis, problemas de apareamiento y técnicas demasiado simples.
- Constatación de diferencias ó asociaciones muy débiles como si fueran mucho mayores.
- Mala interpretación del grado de significación P en los tests y confusión entre significación estadística e importancia clínica

No estamos hablando de una cuestión insignificante, pues las consecuencias del mal uso de la Estadística y de los errores en los principios y métodos estadísticos utilizados en la literatura médica son incalculables desde varios puntos de vista^{5,6,7} como, por ejemplo:

- Los individuos utilizados en la investigación habrán sufrido riesgos ó molestias sin beneficio alguno.
- Los recursos humanos y materiales y el tiempo dedicado al estudio podrían haberse utilizado de forma más productiva en otro proyecto ó en una mejor planificación del mismo.
- A partir de los supuestos hallazgos del estudio, otros pacientes podrían recibir un deficiente tratamiento en el futuro ó sufrir diagnósticos erróneos con riesgos evidentes para su salud.
- El trabajo de otros investigadores se vería afectado al tratar éstos de profundizar en líneas inútiles ó abandonar otras más fértiles.
- Y, lo que es potencialmente más grave, el investigador ó equipo implicado en el trabajo erróneo podría continuar usando metodología estadística inadecuada en el futuro e incitar a otros a hacerlo.



Todo esto refleja la importancia de que el clínico posea los suficientes conocimientos de metodología estadística para desarrollar la facultad crítica y deductiva que necesita para pensar científica, lógica y críticamente sobre los problemas de salud y para comprender la información suministrada por la literatura médica. Todo ello le ayudará en la toma de decisiones racionales relacionadas con la práctica de la asistencia sanitaria y le permitirá leer con espíritu crítico dicha literatura a fin de descubrir si la metodología utilizada puede producir una información útil y detectar tanto afirmaciones erróneas como conclusiones y decisiones que no se deriven de los resultados de un estudio y que carezcan de una base lógica y científica.

Las técnicas y procedimientos estadísticos que pueden haber sido utilizados en cualquier artículo publicado en la literatura médica forman un espectro muy amplio y son de una extensa variedad en cuanto a dificultad y complejidad. No es objetivo de este capítulo, como se comprenderá, hacer una exposición exhaustiva de ellos ni pretender que el lector, al finalizar su lectura, sea un experto en los mismos, sino exponer, de manera breve y concisa, algunos procedimientos estadísticos básicos de gran importancia en la Medicina Basada en la Evidencia y de uso frecuente en la toma de decisiones en la práctica clínica, así como despertar el interés del clínico para que acuda a textos apropiados para aprender ó ampliar sus conocimientos sobre los mismos.

2.1. MUESTREO.

Generalmente, las investigaciones clínicas tratan de sacar conclusiones sobre la efectividad de un tratamiento ó de un procedimiento de diagnóstico, para la generalidad de individuos con unas determinadas características, habitualmente pacientes con una enfermedad concreta, que constituyen lo que, en terminología estadística, se conoce como población. Como es imposible experimentar un nuevo tratamiento en toda esa población de pacientes, se selecciona un grupo de ellos, denominado muestra, en los que se realiza el estudio, para después obtener conclusiones sobre dicho tratamiento para la población origen de esa muestra. El interés de un estudio radica, pues, en la generalización de los resultados: no se está tan interesado en lo que ocurre con los individuos particulares que participan en la investigación como en predecir lo que pueda ocurrir en el futuro con otros individuos similares.

Este proceso, conocido como inferencia estadística, presenta diversos problemas. Uno es el del sesgo de selección⁸, es decir, hasta qué punto los pacientes incluidos en el estudio son similares a aquellos a quienes se quieren aplicar los resultados del estudio o, dicho de otra manera, hasta qué punto son tan distintos como para que los resultados del estudio no sean útiles para tomar decisiones sobre éstos. Realizando la selección de pacientes con las técnicas de muestreo adecuadas⁹, se obtiene una muestra representativa de la población, lo que nos permite garantizar la validez de la inducción de los resultados hacia ésta.

Por otra parte, si las conclusiones que se obtienen para la población se basan en los datos obtenidos en una muestra de la misma, es lógico pensar que, para que esas conclusiones sean fiables, lo primero que hay que exigir es que esa muestra contenga suficiente información de la población. En este sentido, el tamaño de la muestra es uno de los aspectos más importantes de un estudio. Actualmente, no es admisible presentar un protocolo de una investigación médica ó los resultados de la misma,

indicando que se ha usado un número de pacientes concreto sin justificarlo. Es conocido que la calidad científica de los resultados de un estudio, la fiabilidad de las conclusiones ó la representatividad de los parámetros obtenidos están en estrecha relación con el tamaño de las muestras utilizadas. Como norma general, el tamaño de muestra en una investigación debe ser el mínimo necesario, pero a su vez suficiente para proporcionar una respuesta fiable al objetivo final del estudio. La determinación de dicho tamaño mínimo es de naturaleza estadística y depende de varios factores como el tipo de estudio, el tipo de variable utilizada, los riesgos de error que se está dispuesto a asumir, la precisión que se desea obtener, etc.,.

Otro problema es que al estar la muestra de pacientes seleccionada con procedimientos basados en el azar, éste hace que los resultados obtenidos en la misma no coincidan exactamente con los que hubiéramos obtenido si hubiéramos incluido a toda la población de pacientes en el estudio. Se produce, pues, una diferencia entre el valor obtenido para un parámetro en la muestra y el verdadero valor de ese parámetro en la población, un error debido al muestreo. Esta situación provoca que las conclusiones de cualquier estudio para la población no sean fiables al 100%, generándose unos riesgos de error que deben tenerse en cuenta a la hora de la toma de decisiones. Procedimientos estadísticos como la estimación por intervalo de confianza o las pruebas de significación estadística manejan estos inconvenientes y proporcionan soluciones para los mismos.

7

→ 3 INTERVALO DE CONFIANZA.

Aunque existen otros tipos de métodos para la estimación de un parámetro en la población, los procedimientos de estimación por intervalo de confianza son los más utilizados en la literatura médica debido, básicamente, a su fácil interpretación.

Un intervalo de confianza se construye alrededor del valor del parámetro obtenido en la muestra y es un rango de valores, consistente con los datos muestrales, en el que se encuentra, con una cierta seguridad (llamada nivel de confianza), el verdadero valor del parámetro en la población. El nivel de confianza se expresa en términos probabilísticos y se interpreta como la probabilidad de que el verdadero parámetro poblacional esté dentro del intervalo de confianza. La amplitud del intervalo nos informa sobre la precisión de la estimación, comprobándose, por una parte, que cuanto mayor es la muestra y cuanto menor es la dispersión de los datos, más estrecho es el intervalo y, en consecuencia, más precisa es una estimación, y por otra que, cuanto mayor sea el nivel de confianza que utilicemos, peor es la precisión de la estimación. En la investigación médica se utiliza, habitualmente, el nivel de confianza del 95% porque proporciona una suficiente seguridad y no se amplía demasiado el intervalo de confianza. Quizás, una visión frecuentista del nivel de confianza sea más intuitiva para el lector: un nivel de confianza del 95% se puede interpretar como que de cada 100 muestras del mismo tamaño que obtengamos aleatoriamente de la población y para las que calculamos los respectivos intervalos de confianza, es de esperar que en 95 de ellos esté incluido el verdadero valor del parámetro poblacional.

Los métodos estadísticos nos permiten obtener intervalos de confianza para cualesquiera de los parámetros que se usan en la literatura médica (media, proporción, coeficientes de correlación y de regresión, diferencia de proporciones, medidas del efecto de un tratamiento, etc..), pudiendo consultarse en cualquier texto de Bioestadística^{10,11}.

7

→ 4 SIGNIFICACIÓN ESTADÍSTICA.

Un ensayo clínico es de naturaleza básicamente comparativa y, en su versión más simple, consiste en la comparación de un grupo de pacientes a los que se administra un tratamiento experimental (grupo tratamiento), con otro grupo al que se administra un tratamiento control, generalmente un tratamiento estándar ó un placebo (grupo control). Los resultados de uno y otro grupo, lógicamente, no serán exactamente iguales. Se observan, pues, diferencias entre los dos grupos, de mayor ó menor magnitud, que puede ser debidas a algunas de las siguientes causas:

- Los individuos de uno y otro grupo difieren de manera sistemática al inicio del ensayo en factores que pueden influir en el resultado del tratamiento. Se trataría, pues, de grupos no comparables antes de empezar el estudio.
- Los pacientes de los dos grupos difieren durante el desarrollo del ensayo en cuanto a la forma de administración de los tratamientos, manipulación de los pacientes, criterios de valoración, etc., no manteniéndose, pues, la comparabilidad a lo largo del estudio.
- El azar del muestreo al seleccionar los pacientes.
- El verdadero efecto diferente de los dos tratamientos.

La metodología de los ensayos clínicos controlados¹², que está muy protocolizada, nos permite, en teoría, suprimir las dos primeras causas: mediante la aleatorización ó randomización, asignación al azar de los individuos que participan en el ensayo a los grupos de tratamiento, logramos eliminar las diferencias sistemáticas inherentes entre los individuos del grupo experimental y del grupo control al comienzo del estudio y, como consecuencia, se consigue la comparabilidad de los mismos; utilizando procedimientos de enmascaramiento (blinding) como los ensayos a ciego ó a doble ciego, conseguimos mantener la comparabilidad entre los grupos durante el estudio, sobre todo en relación a la administración de los tratamientos, la evolución y seguimiento de los pacientes y la valoración del resultado del tratamiento. Por lo tanto, una vez descartadas las dos primeras causas de esas diferencias observadas, si pudiéramos afirmar sólidamente que es poco probable que las mismas sean explicadas por el azar del muestreo, se podría concluir que las diferencias observadas entre los dos tratamientos son debidas a que uno es más efectivo que el otro.

Las herramientas estadísticas que nos permiten valorar hasta qué punto estas diferencias son debidas al azar de muestreo son las pruebas ó tests de significación estadística, también llamadas contrastes de hipótesis. Son las técnicas estadísticas usadas con más frecuencia en la literatura clínica¹³ y epidemiológica¹⁴ y, a menudo, son mal entendidas y erróneamente interpretadas¹⁵. Se utilizan en todos los campos de la Ciencia, pero de una manera más formal, y en el ámbito particular de un ensayo clínico, en un test de significación estadística se confronta la hipótesis de que los dos tratamientos son igualmente efectivos, llamada hipótesis nula (H0), contra la hipótesis de que tienen distinta efectividad, hipótesis alternativa (H1), y se decide una de las hipótesis mediante la evaluación de la consistencia de los datos obtenidos con la hipótesis nula.

El procedimiento utiliza propiedades estadísticas teóricas (el lector puede consultarlo en cualquier libro de Bioestadística^{10,11}) para obtener un valor p , llamado grado de significación, muy utilizado en la literatura médica, que se interpreta como la probabilidad de equivocarnos si decidiéramos la hipótesis alternativa, o lo que es lo mismo, la probabilidad de que las diferencias observadas entre los dos tratamientos sean debidas sólo al azar del muestreo. De esta manera, el valor p mide la fuerza probatoria de los datos observados contra la hipótesis nula de igual efectividad de los tratamientos, en el sentido de que cuanto menor sea P , mayor es la evidencia contra H_0 . Hoy en día, en la literatura médica, se usa de manera extendida el valor de 0,05 como punto de corte conveniente para la toma de la decisión sobre una u otra hipótesis, aunque la interpretación última la debe hacer el investigador. Así pues, si $p < 0.05$ se entiende que es suficientemente improbable que las diferencias sean debidas sólo al azar, que es poco probable que nos equivoquemos si decidimos H_1 , por lo que existe suficiente evidencia contra la hipótesis nula y, en consecuencia, aceptaremos la hipótesis alternativa de que un tratamiento es más efectivo que otro con ese riesgo p de error; en este caso se dice que la diferencia entre los tratamientos es estadísticamente significativa. Por el contrario, si $p > 0.05$ se considera que ya es demasiado probable que las diferencias sean debidas al azar del muestreo, existe demasiada probabilidad de equivocarse si decidimos H_1 , con lo que la evidencia contra la hipótesis nula es débil y los datos son consistentes con la misma; en esta situación decidiremos la hipótesis nula y diremos que la diferencia entre los tratamientos no es significativa. Se debe tener cuidado cuando se decide que la diferencia no es significativa pues, en ese caso, debemos entender que no hemos podido demostrar que haya diferencia entre los tratamientos en la población en cuanto a su efectividad, pero no hemos probado que sean iguales; dicha diferencia aún puede existir y puede haber ocurrido que la muestra no haya sido lo suficientemente grande como para detectarla.

Existen muchos tipos de contrastes de hipótesis para comparar dos tratamientos^{10,11} que van desde los tests t de Student de comparación de medias y los tests no paramétricos si el criterio de valoración del resultado del tratamiento es cuantitativo, a los tests χ^2 de comparación de proporciones si dicho criterio es cualitativo, pasando por los tests referidos a las medidas del efecto de los tratamientos que veremos más adelante.

4.1. SIGNIFICACIÓN ESTADÍSTICA Y SIGNIFICACIÓN CLÍNICA.

Uno de los aspectos más importantes que debe tener en cuenta el médico a la hora de la toma de decisiones clínicas a partir de los resultados publicados en la literatura médica, es no confundir significación estadística con significación clínica¹⁶. Mientras que la significación estadística, como hemos dicho, evalúa hasta qué punto los datos sugieren que las diferencias observadas entre los efectos de los tratamientos pueden ser explicadas ó no por la casualidad, la significación clínica se refiere a la magnitud y la importancia de esas diferencias en la práctica clínica¹⁷. La interpretación de los resultados de un estudio no debe basarse solamente en el valor del grado de significación p , ya que no mide la importancia clínica de la diferencia observada y, sin embargo, se ve muy influido por el tamaño muestral¹⁸. De hecho, puede ocurrir que una diferencia muy pequeña, sin relevancia clínica, puede haberse detectado como significativa debido a que se ha estudiado un elevado número de pacientes. Y viceversa, una diferencia grande entre los grupos, que parece clínicamente importante, puede que no se detecte como significativa por disponer de pocos pacientes en el estudio ó por existir una gran dispersión en los datos de los pacientes.

Una manera de considerar la significación clínica de la diferencia del efecto de los tratamientos a la hora de realizar un test de significación estadística, ya que éste no nos informa sobre cuanto mejor es un tratamiento que otro, podría ser combinarlo con el intervalo de confianza de dicha diferencia. Los intervalos de confianza identifican el rango donde puede estar, con alta probabilidad, el verdadero efecto del tratamiento. De hecho, hay autores^{19,20} que piensan que, en la toma de decisiones en la clínica, los intervalos de confianza son más útiles para los clínicos que los tests de significación estadística, proponiendo realizar éstos últimos mediante la utilización, de forma complementaria, de los intervalos de confianza, de manera que si el valor de la hipótesis nula (que suele ser diferencia cero entre medias ó entre proporciones) no está incluido en el intervalo de confianza al nivel de confianza del 95%, se rechaza la hipótesis nula y se deciden diferencias significativas al nivel de significación del 5%. No obstante, en este procedimiento existe el inconveniente de que no se puede calcular el verdadero grado de significación p pero, sin embargo, nos permite disponer adicionalmente de un rango de valores del efecto del tratamiento.

7

→ **5 MEDIDAS DEL EFECTO DEL TRATAMIENTO.**

El énfasis que se viene haciendo en la importancia de incorporar de forma crítica los resultados de la investigación científica en la práctica médica habitual, conlleva la necesidad de expresar los resultados de forma clara, inteligible y fácilmente trasladable a las situaciones que el médico encuentra en su actividad diaria. En este sentido, para los clínicos que buscan aplicar los resultados de los ensayos clínicos ó de las revisiones sistemáticas a la atención sanitaria de sus pacientes, es importante valorar de forma adecuada la significación clínica de la diferencia del efecto de dos tratamientos, sea ésta ó no estadísticamente significativa, utilizando medidas específicas del efecto que cuantifiquen la mejoría de un tratamiento sobre el otro y que sean más cercanas al tipo de razonamiento que se usa en la toma de decisiones clínicas.

En la literatura clínica y epidemiológica se usan, para ello, distintos índices²¹ y algunos trabajos indican que, dependiendo de qué medidas del efecto se empleen en un determinado estudio, el impacto de una intervención ó un tratamiento puede parecer grande ó pequeño²² ó que el grado en el que los médicos están dispuestos a prescribir un tratamiento depende, en gran medida, de qué índice se use para representar su eficacia^{23,24}. Todos esos índices son correctos y legítimos para presentar los resultados, informando de facetas diversas del efecto de los tratamientos, por lo que conviene aquí definir y comentar brevemente los más utilizados.

Las medidas del efecto del tratamiento que estudiaremos tienen un elemento de partida común: el criterio de valoración del resultado del tratamiento en cada paciente es una variable binaria (variable del tipo sí o no) que representa un evento que puede ocurrir o no ocurrir como, por ejemplo, muerte, curación, recidiva, mejoría, ocurrencia de una enfermedad, etc.. Cuando ese resultado es una variable cuantitativa, generalmente continua, como por ejemplo, tensión arterial, glucemia, etc., lo habitual es transformarlas en binarias: hipertensión o normotensión, hiperglucemia o normoglucemia, etc.

En un ensayo clínico, el resultado observable en cada paciente es que dicho paciente presente ó no el evento en estudio; el resultado para un grupo de pacientes se resume mediante la proporción de pacientes en los que el evento ocurre. Esta proporción estima la probabilidad o “riesgo” del evento en ese grupo. Este último término es el utilizado de manera habitual, aunque genera cierta confusión terminológica porque, en ensayos clínicos, se puede estudiar unas veces eventos adversos (muerte, enfermedad, efectos secundarios, etc,..) y otras veces eventos beneficiosos (supervivencia, curación, mejoría, etc,..) y para ambos se usa el término “riesgo”, que hace pensar sólo en eventos adversos.

En cualquier ensayo clínico, los resultados del mismo se pueden expresar en una tabla como la siguiente:

	Evento	No evento	
Tratamiento	<i>a</i>	<i>b</i>	<i>a + b</i>
Control	<i>c</i>	<i>d</i>	<i>c + d</i>
	<i>a + c</i>	<i>b + d</i>	

En esta situación, la estimación de los riesgos en el grupo tratamiento (R_t) y en el grupo control (R_c) sería:

$$R_t = \frac{a}{a+b} \qquad R_c = \frac{c}{c+d}$$

A partir de aquí, para evaluar el efecto de un tratamiento hay que comparar el riesgo en el grupo tratamiento con el riesgo en el grupo control; las distintas formas de compararlos originan las diversas medidas del efecto del tratamiento que vamos a considerar.

5.1. REDUCCIÓN ABSOLUTA DEL RIESGO.

La manera más sencilla de comparar dos riesgos, en definitiva dos números, es hallando su diferencia. La reducción absoluta del riesgo (RAR) se define como la diferencia del riesgo del grupo tratamiento y el riesgo del grupo control:

$$RAR = R_t - R_c = \frac{a}{a+b} - \frac{c}{c+d}$$

Como ejemplo, consideremos la siguiente situación en la que el evento de interés es la muerte:

	Muertos	Vivos	
Tratamiento	18	46	64
Control	29	36	65
	47	82	129

Los riesgos en los grupos tratamiento y control serían

$$R_t = \frac{18}{64} = 0.2813 \quad (28.13\%) \quad R_c = \frac{29}{65} = 0.4462 \quad (44.62\%)$$

Y la reducción absoluta del riesgo

$$RAR = 0.4462 - 0.2813 = 0.1649 \quad (16.49\%)$$

La RAR se interpreta como el porcentaje de pacientes que han evitado el evento adverso como consecuencia de recibir el tratamiento experimental en lugar del control. En nuestro ejemplo, el RAR indica que el tratamiento experimental reduce el riesgo de muerte en 0,1649 o un 16.49%. Dicho de otra manera, por cada 100 pacientes que sigan el tratamiento experimental se evitarán entre 16 y 17 muertes respecto a las que se hubieran producido si se hubiera empleado el tratamiento control. Por tanto, es un índice que expresa las consecuencias de dar el tratamiento, por ello a veces se le denomina también “reducción atribuible del riesgo”, o abreviadamente “riesgo atribuible”. También se puede expresar como “diferencia absoluta de riesgo”. Su principal inconveniente es que se expresa con un número pequeño, que quizás explique por qué es el índice con el que los médicos perciben que el efecto es menor²².

La ausencia total de efecto del tratamiento sería una $RAR = 0$, por lo que se puede realizar un test de significación estadística para la reducción absoluta del riesgo contrastando la hipótesis nula $H_0: RAR = 0$ contra la hipótesis alternativa $H_1: RAR \neq 0$.

Se puede obtener el intervalo de confianza para la RAR teniendo en cuenta que ésta no es más que una diferencia de riesgos que se estiman mediante proporciones. De este modo, cuando las muestras son suficientemente grandes, el intervalo de confianza de la reducción absoluta del riesgo, al nivel de confianza del 95%, se puede calcular a través de la expresión

$$RAR \pm 1.96 \cdot \sqrt{\frac{R_t \cdot (1 - R_t)}{n_t} + \frac{R_c \cdot (1 - R_c)}{n_c}}$$

siendo n_t y n_c el número de pacientes de los grupos tratamiento y control respectivamente.

5.2. RIESGO RELATIVO.

Otra forma de comparar dos riesgos es a través de su cociente. El riesgo relativo (RR) es el cociente entre el riesgo del grupo tratamiento y el riesgo del grupo control:

$$RR = \frac{R_t}{R_c} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

Observando esta definición, el riesgo relativo se puede interpretar como el número de veces que es más probable que presente el evento un paciente del grupo tratamiento que otro del grupo control y se deduce que si $RR < 1$ el tratamiento experimental protege de la aparición del evento, si $RR = 1$ resulta indiferente y si $RR > 1$ favorece la aparición del mismo.

En nuestro ejemplo, el riesgo relativo sería

$$RR = \frac{18/64}{29/65} = \frac{0.2813}{0.4462} = 0.63$$

por lo que podemos decir que el riesgo en el grupo tratamiento es 0.63 veces el riesgo en el grupo control, es decir, un 63% del mismo.

Para obtener el intervalo de confianza para el riesgo relativo, dado que éste tiene una distribución muy asimétrica, se debe utilizar el $\ln(RR)$, que sigue aproximadamente una distribución normal, obteniéndose la siguiente expresión como intervalo de confianza para el riesgo relativo RR al nivel de confianza del 95%:

$$e^{\ln(RR) \pm 1.96 \times \sqrt{\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}}}$$

En nuestro ejemplo,

$$e^{\ln(0.63) \pm 1.96 \times \sqrt{\frac{1}{18} - \frac{1}{64} + \frac{1}{29} - \frac{1}{65}}}$$

lo que da un intervalo de confianza para el riesgo relativo de (0.39 ; 1.01)

Como hemos visto, cuando los tratamientos experimental y control tengan el mismo efecto, el riesgo relativo sería $RR = 1$, por lo que un test de significación estadística para el riesgo relativo consistiría en el contraste de la hipótesis nula $H_0: RR = 1$ contra la hipótesis alternativa $H_1: RR \neq 1$. Aunque existen otras formas de llevarlo a cabo, dicho test se podría realizar al nivel de significación del 5%, tal y como se ha señalado con anterioridad en este capítulo, comprobando si el valor de esta hipótesis, es decir, $RR = 1$, está ó no dentro de este intervalo de confianza.

5.3. REDUCCIÓN RELATIVA DEL RIESGO.

La reducción relativa del riesgo (RRR) se define como el cociente entre la reducción absoluta del riesgo RAR y el riesgo en el grupo control R_c .

$$RRR = \frac{RAR}{R_c} = \frac{R_t - R_c}{R_c}$$

Considerando el riesgo en el grupo control como riesgo basal, es decir, riesgo sin tratamiento ó con el tratamiento estándar, la reducción relativa del riesgo se interpreta como la proporción ó el porcentaje del riesgo basal que es eliminado como consecuencia de la nueva terapia.

Si el evento de interés es un evento adverso, el riesgo relativo será mayor que 1 y la reducción relativa del riesgo se puede expresar como

$$RRR = \frac{R_c - R_t}{R_c} = 1 - RR$$

mientras que si es un evento beneficioso, el riesgo relativo es menor que 1 y la reducción relativa del riesgo sería

$$RRR = \frac{R_t - R_c}{R_c} = RR - 1$$

observándose que la RRR es la parte del riesgo relativo que falta para 1 ó que excede de 1, respectivamente.

En nuestro ejemplo,

$$RRR = \frac{RAR}{R_c} = \frac{0.1649}{0.4462} = 0.3696 \quad (36.96\%)$$

es decir, la administración del nuevo tratamiento supone una reducción del 36.96% del riesgo del grupo control.

A la reducción relativa del riesgo se le llama también “fracción atribuible” ó “diferencia relativa del riesgo” e ilustra el beneficio del tratamiento en términos relativos. Esto es una ventaja pero también un inconveniente, pues al dejar de considerar la magnitud del riesgo en el grupo control (riesgo basal) se pueden obtener sobrestimaciones o subestimaciones del impacto del tratamiento si el riesgo control fuera muy bajo o muy alto, respectivamente. Así, por ejemplo, si el riesgo de eventos adversos sin tratamiento es bajo en un determinado grupo de pacientes al comienzo del estudio, una pequeña reducción de ese riesgo en términos absolutos da lugar a la misma RRR que una reducción mayor en un grupo de pacientes que tienen un riesgo basal también mayor. Es más, existe fundada evidencia^{25,26} de que mientras la reducción absoluta del riesgo (RAR) disminuye cuanto menor es el riesgo basal, la reducción relativa del riesgo, a menudo, permanece constante.

5.4. NÚMERO NECESARIO DE PACIENTES PARA TRATAR.

El número necesario de pacientes para tratar (NNT) se define como número de pacientes que se estima que es necesario que reciban el tratamiento experimental en lugar del control para que un paciente adicional evite el evento adverso (u obtenga el beneficio si se trata de un evento beneficioso).

Para deducir la expresión para su cálculo, utilicemos nuestro ejemplo: en el apartado 5.1 obtuvimos una reducción absoluta del riesgo de $RAR = 0.1649$ (16.49%), lo que quiere decir que por cada 100 pacientes que sigan el tratamiento experimental en lugar del control, se evitarán 16.49 muertes. Entonces, para evitar una muerte, y mediante una sencilla regla de tres, se necesitará tratar un número de pacientes

$$NNT = \frac{1}{0.1649} = 6.06$$

es decir, que se espera que se evite una muerte por cada, aproximadamente, 6 pacientes tratados con la nueva terapia.

Así pues, el número necesario de pacientes para tratar NNT se puede calcular mediante del inverso de la reducción absoluta del riesgo RAR, es decir,

$$NNT = \frac{1}{RAR} \quad \text{ó} \quad NNT = \frac{100}{RAR(\%)}$$

según que la RAR esté expresada en tanto por uno ó en tanto por ciento respectivamente.

Evidentemente, cuanto menor sea el NNT más importante es el efecto del tratamiento y viceversa, a un menor efecto del tratamiento le corresponde un mayor NNT, llegando a la teórica situación límite de que cuando el efecto del tratamiento es nulo ($RAR=0$), el NNT tendría valor infinito.

El NNT se ha popularizado bastante en los últimos años a raíz del auge de la Medicina basada en la evidencia, ya que es una medida clínica muy útil de las consecuencias de un nuevo tratamiento, pues expresa el efecto del mismo en unos términos intuitivos, más próximos a su significado clínico, que permiten fácilmente comparar las ventajas del tratamiento con sus inconvenientes, acercándose, pues, al modo de razonamiento que se usa en la toma de decisiones clínicas. Es una medida concisa, fácil de calcular, fácil de comprender y tiene la ventaja de que no sólo proporciona información estadística, sino también información clínica. Ha sido considerado como la medida del efecto más representativa de la significación clínica²⁷.

El NNT es un parámetro eficaz por su capacidad para comunicar y presentar el impacto de un tratamiento o de una intervención, pero por ello mismo, debido a su capacidad de simplificación, hay que ser cautos a la hora de utilizarlo e interpretarlo, teniendo en cuenta que existen otras alternativas. En este sentido, como cualquier resultado estadístico, el NNT se puede acompañar con su intervalo de confianza, aunque el cálculo de éste es un poco problemático y su interpretación resulta un tanto confusa si no hay un efecto claro del tratamiento, es decir, si el tratamiento experimental no es significativamente más eficaz ó, incluso, si es menos beneficioso que el control^{28,29}.

Por último, como consecuencia de su reciente popularidad, el NNT también se tiende a utilizarlo, de manera indiscriminada, como medida global en los meta-análisis de ensayos clínicos. Pero hay que tener mucha precaución a la hora de interpretar los NNT obtenidos a partir de un meta-análisis³⁰, pues se debe siempre analizar el riesgo base en cada uno de los estudios individuales, ya que si hay heterogeneidad en dichos riesgos basales, es decir, si éstos son diferentes, su utilidad es más que dudosa y el NNT global puede resultar poco informativo. En esos casos es preferible utilizar un índice relativo como el riesgo relativo ó el odds ratio, pues, aunque no nos den una medida del impacto real del tratamiento como el RAR y el NNT, se ha evidenciado que los índices relativos tienden a ser más homogéneos entre diferentes ensayos del mismo tratamiento y están menos influenciados por el riesgo basal³¹.

5.5. NÚMERO NECESARIO DE PACIENTES PARA PERJUDICAR.

De igual modo que el NNT, se puede usar el número necesario para perjudicar para evaluar efectos adversos del tratamiento. Este índice, representado por las siglas de su nombre en inglés NNH (number needed to harm), se define como el número de pacientes que deberían recibir el tratamiento experimental en lugar del control para que un paciente adicional obtenga el perjuicio. Sólo tiene sentido emplearlo cuando el riesgo del evento perjudicial sea mayor en el grupo tratamiento que en el grupo control, como, por ejemplo, cuando los efectos secundarios son más frecuentes en el tratamiento experimental que en el control.

Su cálculo es similar al del NNT, es decir,

$$NNH = \frac{1}{RAR} \quad \text{ó} \quad NNH = \frac{100}{RAR (\%)}$$

y todas las demás consideraciones realizadas para aquél en el apartado anterior son válidas para el NNH, con la única diferencia, como hemos indicado, de que estamos tratando con efectos adversos del tratamiento.

La toma en consideración del NNT y del NNH conjuntamente permite al médico tener en cuenta, al mismo tiempo, las ventajas e inconvenientes de un tratamiento, tal y como se hace en la práctica clínica, ayudándole en la toma de decisiones con argumentos más sólidos.

5.6. ODDS RATIO.

Esta es una medida de asociación entre variables binarias de uso bastante frecuente en la literatura clínica y epidemiológica que, bajo ciertas condiciones, también se utiliza como medida del efecto de un tratamiento.

El odds es un término para el que en castellano no existe una traducción comúnmente aceptada, empleándose con cierta frecuencia las palabras “posibilidad” ó “ventaja”. En el ámbito estadístico, es otra forma de presentar un riesgo, un poco distinta a la proporción ó al porcentaje que hemos visto anteriormente. El odds es el cociente entre la probabilidad de que el episodio de interés ocurra y la probabilidad de que no ocurra y, habitualmente, se estima por el cociente entre el número de veces que ha ocurrido el evento y el número de veces que no ha ocurrido, midiendo, por tanto, las veces que es más probable que ocurra el evento que las que no ocurra. Así, un odds de 4 indica que por cada 4 pacientes que presenten el evento, habrá 1 que no lo presentará.

Con nuestra notación, los odds en uno y otro tratamiento se obtendrían como

$$\text{Odds}_t = \frac{a}{b} \quad \text{y} \quad \text{Odds}_c = \frac{c}{d}$$

En nuestro ejemplo, el odds en el grupo tratamiento sería 18/46 mientras que en el grupo control 29/36; ambos son menores que 1, lo que quiere decir que, en ambos grupos, es menos frecuente la muerte del paciente que la supervivencia.

Si queremos comparar esta medida especial del riesgo en cada grupo, el odds, entre los dos tratamientos, podemos calcular el cociente entre ambos odds obteniendo, así, el conocido odds ratio (OR)

$$\text{OR} = \frac{\text{Odds}_t}{\text{Odds}_c} = \frac{\cancel{a}/\cancel{b}}{\cancel{c}/\cancel{d}} = \frac{a \cdot d}{b \cdot c}$$

En caso de que no haya efecto del tratamiento, el riesgo ó probabilidad de que ocurra el evento es el mismo para ambos grupos, por lo que los odds también coincidirán y, por tanto, el odds ratio será $OR = 1$. Por otra parte, si el $OR > 1$ se interpretará como que el tratamiento experimental favorece la aparición del evento pues su odds es mayor que el del control y viceversa, si el $OR < 1$ diremos que protege de la aparición del mismo. Además, cuanto mayor sea el efecto del tratamiento, en uno u otro sentido, más distintos son los odds y, por tanto, más se alejará el OR de 1 (en el sentido de mayor o menor que 1). En este aspecto, el odds ratio tiene un comportamiento similar al riesgo relativo.

Con los datos de nuestro ejemplo, el odds ratio sería

$$OR = \frac{18 \cdot 36}{46 \cdot 29} = 0.486$$

lo que quiere decir que, en nuestros datos, el tratamiento experimental reduce la aparición del evento muerte respecto al control.

Al igual que con el riesgo relativo, se podría obtener el intervalo de confianza para el odds ratio utilizando su logaritmo, pues el OR tiene una distribución muy asimétrica y el $\ln(OR)$ sigue aproximadamente una distribución normal, obteniéndose la siguiente expresión como intervalo de confianza para el odds ratio OR al nivel de confianza del 95%:

$$e^{\ln(OR) \pm 1.96 \times \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}}$$

En nuestro ejemplo, este intervalo de confianza sería

$$e^{\ln(0.486) \pm 1.96 \times \sqrt{\frac{1}{18} + \frac{1}{46} + \frac{1}{29} + \frac{1}{36}}}$$

lo que da un intervalo de confianza para el odds ratio de (0.23 ; 1.01)

Por otra parte, al igual que con el riesgo relativo, como cuando los tratamientos experimental y control tienen el mismo efecto, el odds ratio es $OR = 1$, se podría realizar un test de significación estadística para el odds ratio que consistiría en el contraste de la hipótesis nula $H_0: OR = 1$ contra la hipótesis alternativa $H_1: OR \neq 1$. Del mismo modo, dicho test se podría realizar al nivel de significación del 5%, comprobando si el valor de la hipótesis nula $OR = 1$ está ó no dentro de este intervalo de confianza.

Hemos visto, pues, que el odds ratio se comporta de manera parecida al riesgo relativo, sin embargo tiene una interpretación mucho menos intuitiva que éste, por lo que el OR ha recibido frecuentes críticas por lo proclive que resulta a interpretaciones erróneas³². Pero a pesar de ser más simple el RR que el OR, ¿cuáles son las razones por las que es tan popular éste último y se usa con tanta frecuencia en la literatura médica?

En primer lugar, porque se puede usar en cualquier tipo de estudio que genere una tabla ²², como los estudios observacionales prospectivos, retrospectivos ó transversales, ó los estudios experimentales, como los ensayos clínicos. En particular, en los estudios de casos y controles, mientras que sí se puede utilizar el OR para analizar la asociación entre la exposición a un factor y un evento determinado, no es posible estimar el RR ya que, en este tipo de estudios, la distribución de casos y controles en la muestra se elige a conveniencia en el diseño del estudio, por lo que no representa la distribución de la población y, en consecuencia, no se pueden conocer los riesgos en expuestos y no expuestos en la población.

Por otra parte, cuando el riesgo, ó el odds, es muy pequeño en ambos grupos (hecho muy frecuente en estudios observacionales pero no en ensayos clínicos), el OR y el RR son muy parecidos, pudiendo considerarse el primero como una buena aproximación del segundo. Así, en aquellos ensayos clínicos que presenten riesgos suficientemente bajos, por ejemplo, inferiores al 20%, el OR se interpretaría igual que el riesgo relativo, es decir, número de veces que es más probable que presente el evento un paciente del grupo tratamiento que otro del control, debiendo evitarse el uso del OR cuando los riesgos no son demasiado bajos³³.

Otras razones son de tipo matemático: el odds ratio no varía aunque se cambie el orden de las categorías en cualquiera de las variables, lo que no ocurre con el riesgo relativo. Por otra parte, en un análisis de regresión logística, el OR permite examinar el efecto que otras variables pueden causar en una asociación (como podrían ser por ejemplo la edad, el sexo, etc.), ajustando la relación en estudio por el efecto de dichas variables, interpretándose la exponencial de un coeficiente del modelo de regresión logística como el odds ratio de la variable correspondiente.

1. Aalen OO. Medical Statistics: no time for complacency. *Statistical Methods in Medical Research* 2000; 9: 31-40.
2. Altman DG. Statistics in Medical Journals. *Statistics in Medicine* 1982; 1: 59-71.
3. Altman DG. Statistics in medical journals: development in the 1980's. *Statistics in Medicine* 1991; 10: 1897-1913.
4. Altman DG, Bland JM. Improving doctors' understanding of statistics (with discussion). *Journal of the Royal Statistical Society A* 1991; 154: 223-267.
5. Altman DG. The scandal of poor medical research. *British Medical Journal* 1994; 308: 283-284.
6. Altman DG. Statistical reviews for medical journals. *Statistics in Medicine* 1998; 17: 2661-2674.
7. Altman DG. Statistics in medical journals: some recent trends. *Statistics in Medicine* 2000; 19: 3257-3289.
8. Ellenberg JH. Selection bias in observational and experimental studies. *Statistics in Medicine* 1994;13:557-67.
9. Silva Aycaguer LC. Muestreo para la investigación en ciencias de la salud. Madrid: Díaz de Santos, D.L., 1993.
10. Martín Andrés A, Luna del Castillo JD. Bioestadística para las Ciencias de la Salud. 4ª ed. Madrid: Ed. Norma; 1994.
11. Armitage P, Berry G. Estadística para la Investigación Biomédica. Barcelona: Doyma; 1992.
12. Jadad AR. Randomised controlled trials. London: BM J Books, 1998.
13. Guyatt G, Jaeschke R, Heddle N, Cook D, Shannon H, Wermuth L. Basic statistics for clinicians: 1. Hypothesis testing. *CMAJ* 1995; 152: 27-32.

14. Poole C. Low P-values or narrow confidence intervals: which are more durable?. *Epidemiology* 2001; 12: 291-4.
15. Sterne JAC, Smith GD. Sifting the evidence - what's wrong with significance tests? *British Medical Journal* 2001; 322: 226-31.
16. Porta M, Plasencia A, Sanz F. La calidad de la información clínica (III): ¿estadísticamente significativo o clínicamente importante?. *Medicina Clínica* 1988; 90: 463-68.
17. Leung WC. Balancing statistical and clinical significance in evaluating treatment effects. *Postgrad Med J* 2001; 77: 201-204.
18. Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. *Ann Intern Med* 1999; 130:995-1004.
19. Montori VM, Kleinbart J, Newman TB, Keitz S, et al. Tips for learners of evidence-based medicine: 2. Measures of precision (confidence intervals). *CMAJ* 2004; 171(6): 611-615.
20. Feinstein AR. P-values and confidence intervals: two sides of the same unsatisfactory coin. *J Clin Epidemiol* 1998; 51:355-60.
21. Fahey T, Griffiths S, Peters TJ. Evidence based purchasing: understanding results of clinical trials and systematic reviews. *British Medical Journal* 1995; 311: 1056-1060.
22. Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness?. *Ann Intern Med* 1992; 117: 916-921.
23. Bobbio M, Demichelis B, Giustetto G. Completeness of reporting trial results: effect on physicians' willingness to prescribe. *Lancet* 1994; 343: 1209-1211.
24. Meneu Ricardo R, Peiró S, Márquez Calderón S. Influencia de la presentación de los resultados de los ensayos clínicos en la intención de prescribir: relativizando el riesgo relativo. *Aten Primaria* 1998; 21: 446-450.
25. Barratt A, Wyer PC, Hatala R, McGinn T, Dans AL, Keitz S, et al. Tips for learners of evidence-based medicine: 1. Relative risk reduction, absolute risk reduction and number needed to treat. *CMAJ* 2004;171(4): 353-358.



26. Schmid CH, Lau J, McIntosh MW, Cappelleri JC. An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. *Statistics in Medicine* 1998;17:1923-42.

27. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *British Medical Journal* 1995; 310: 452-454.

28. Altman DG. Confidence intervals for the number needed to treat. *British Medical Journal* 1998; 317: 1309-1312.

29. Walter SD. Number needed to treat (NNT): estimation of a measure of clinical benefit. *Statistics in Medicine* 2001; 20(24): 3947-3962.

30. Smeeth L, Haines A, Ebrahim S. Numbers needed to treat derived from meta-analyses-sometimes informative, usually misleading. *British Medical Journal* 1999; 318: 1548-1551.

31. Deeks JJ, Altman DG, Dooley G, Sackett DL. Choosing an appropriate dichotomous effect measure for meta-analysis: empirical evidence of the appropriateness of the odds ratio and relative risk. *Controlled Clinical Trials* 1997; 18: S84-S85.

32. Sackett DL, Deeks JJ, Altman DG. Down with odds ratios!. *Evidence-Based Med* 1996; 1: 164-166.

33. Altman DG, Deeks JJ, Sackett DL. Odds ratios should be avoided when events are common [letter]. *British Medical Journal* 1998; 317: 1318.



Acotado Automático de Términos: Brevemente, consiste en el enlace automático de una palabra o frase (que no es un término MeSH) al término MeSH que mejor lo describe. Una descripción más completa en el apartado correspondiente. Sinónimos (Sin): Automatic Term Mapping, ATM, Mapeo automático de términos.

Aleatorización: En un estudio en el que se comparan dos o más modalidades de tratamiento, es conveniente que los grupos formados sean parecidos en todas las características pronósticas excepto en el tratamiento recibido, de modo que cualquier diferencia en el curso clínico que se registre podrá ser atribuida a los diferentes tratamientos administrados (y sólo a ellos). La asignación aleatoria consiste en distribuir a cada participante de uno de los grupos de tratamiento por un método disciplinado de azar, de modo que cada sujeto tenga exactamente las mismas probabilidades de formar parte de uno u otro grupo de tratamiento. Sin: Randomización.

Alertas: Método usado para la difusión de información mediante mensajes electrónicos a ordenadores, agendas electrónicas e incluso teléfonos móviles. Representa una forma de actualización rápida y eficaz.

Análisis coste/beneficio: Medición de los costes y beneficios de una intervención en términos comensurables, generalmente monetarios. Permite determinar en primer lugar si una intervención concreta ofrece un beneficio global neto de bienestar, y en segundo lugar cómo se compara la ganancia de bienestar obtenida con esta intervención, con la obtenida con intervenciones alternativas. Al valorar todos los costes y beneficios en las mismas unidades, el análisis coste-beneficio compara diversas intervenciones con el criterio del beneficio neto (véase). Análisis económico en el que se consideran los costes de la asistencia sanitaria y la pérdida de ganancias netas debida a muerte o incapacidad, traduciendo todos los costes y beneficios a unidades monetarias.

Análisis coste/efectividad: Modalidad de estudio económico en la que las consecuencias de diferentes intervenciones pueden ser diferentes pero pueden ser medidas en unidades naturales idénticas, Mide los beneficios sanitarios en términos de unidades naturales, como años de vida salvados o mejorías del estado funcional. Dado que los costes no se expresan en las mismas unidades que los beneficios, su razón (ratio coste-efectividad) constituye una herramienta que permite valorar la eficiencia relativa. Análisis económico que compara los costes médicos incrementales y los resultados sobre la salud de programas sanitarios alternativos. Al contrario que el análisis de coste beneficio, los efectos sobre la salud se expresan en diversas unidades, como vida-años ganados, días libres de síntomas, casos evitados, etc., en lugar de en unidades monetarias

Análisis coste-utilidad: Adaptación del análisis coste-efectividad que mide el efecto de una intervención a la vez sobre los aspectos cuantitativos y cualitativos de la salud, mediante el uso de una variable denominada AVAC -Años de Vida Ajustados por Calidad- o QALYs . También evalúa la eficiencia relativa con una razón incremental, que en este caso se denomina ratio coste-utilidad.



Análisis de decisiones (AD): Es la aplicación de métodos explícitos y cuantitativos que cuantifican e integran los efectos del tratamiento, el pronóstico y los valores del paciente para analizar una decisión en condiciones de incertidumbre.

Análisis de minimización de costes: Modalidad de análisis coste-efectividad donde los resultados o beneficios de las alternativas se considera que son equivalentes o intercambiables para el mismo tipo de pacientes. A igualdad de resultados, la atención de los investigadores se centra en la diferencia de costes asociados.

Análisis de sensibilidad: En economía de la salud, técnica por la que se repiten las comparaciones entre recursos necesarios y resultados, pero con sucesivas modificaciones de las suposiciones en las que se basan los cálculos. De este modo, el análisis de sensibilidad pone a prueba la robustez de las conclusiones mediante variaciones de los ingredientes sobre los que existe incertidumbre.

Análisis de regresión: En un conjunto de datos sobre la variable dependiente y sobre una o más variables independientes, x_1, x_2, \dots, x_i , consiste en determinar el modelo matemático más ajustado que describa y como una función de las x o para predecir y a partir de las x . Término general para los métodos de análisis que se ocupan de estimar los parámetros de alguna posible relación entre una variable respuesta y una o más variables explicativas

Análisis de la varianza (ANOVA): Conjunto de técnicas estadísticas para conocer el modo en que el valor medio de una variable es afectado por diferentes tipos de clasificaciones de los datos. Con el análisis de la varianza se pueden ajustar las estimaciones del efecto de un tratamiento según otros factores como sexo, edad, gravedad, etc.

Análisis estratificado: Análisis de los resultados de un ensayo clínico o un estudio observacional, con el fin de examinar si hay diferencias de respuestas asociadas a alguna característica (por ej., edad o sexo).

Análisis factorial: Procedimiento estadístico empleado para la validación de constructo que permite identificar grupos o conglomerados de variables relacionadas.

Análisis multivariante: Cualquier método analítico que permite el estudio simultáneo de dos o más variables dependientes. Término genérico para los métodos que están diseñados para determinar las contribuciones relativas de diferentes causas a un solo suceso o resultado

Análisis por intención de tratar: En un ensayo clínico, análisis de los datos según el tratamiento asignado inicialmente por distribución aleatoria, independientemente de que el paciente no recibiera después este tratamiento.

Árbol de decisiones: Expresión gráfica de un análisis de decisiones.

Artículo científico: informe escrito que comunica por primera vez los resultados de una investigación. Los artículos científicos publicados en miles de revistas científicas componen la literatura primaria de la ciencia

Artículo de revisión: artículos de síntesis que resumen el conocimiento de un tema determinado y durante un periodo de tiempo definido. Junto con los libros, los artículos de revisión constituyen la literatura científica secundaria.

Asiento: Conjunto de información que, en una base de datos, se refiere a un mismo elemento. En Bases de Datos de bibliografía, es el conjunto de información que se refiere a una misma publicación. Sin: Registro, Record.

ATM: Ver Acotado Automático de Términos.

Atributos: Propiedades o características del instrumento que merecen consideraciones separadas e independientes en la evaluación.

Auditoria: Examen o estudio que evalúa e informa sobre la medida en que una condición, proceso o desempeño se ajusta a estándares o criterios predeterminados.

Automatic explosion: Ver Explode.

Automatic Term Mapping: Ver Acotado Automático de Términos.

AVAC: Índice que combina la supervivencia con la calidad de vida. La medida de la calidad de vida no está estandarizada y suele variar de un estudio a otro, según la enfermedad, el tratamiento evaluado y las preferencias de los autores. Para calcular un AVAC se multiplica la duración del estado de salud (en años) por un factor que representa la calidad (“utilidad”) de este estado de salud. Sin: QALY (Quality Adjusted Life Years).

Azar: Se denomina así al proceso que ocurre debido a la casualidad, de forma imprevisible, sin planeamiento ni influencia consciente.

Base de datos: Conjunto de datos homogéneos y ordenados, que se presentan normalmente, de forma legible, por ordenador y que se refieren a una materia determinada. Colección estructurada de datos que está organizada de tal manera que se puede acceder a ella con una amplia variedad de programas de aplicación.

Beneficio neto: Diferencia entre los costes y los beneficios de una intervención.

Bioestadística: Aplicación de la teoría y los métodos de la estadística a las ciencias biológicas. En sentido estricto, rama de la ciencia que aplica los métodos estadísticos a los problemas biológicos, aunque ahora se utiliza la mayoría de las veces para referirse a la estadística aplicada a la medicina y las ciencias de la salud.

Blinding: Ver Enmascaramiento.

Buscadores: Motores de búsqueda de información en Internet que identifican la información contenida en bases de datos formadas a partir de publicaciones previamente indexada. Ver Indización.

Calidad de Vida Relacionada con la Salud (CVRS): Impacto que una enfermedad y el tratamiento consiguiente tienen sobre la percepción del paciente de su satisfacción y bienestar físico, psíquico, social y espiritual.

Campo: Cada una de las unidades de información en que se divide un asiento. Sin: Field.

Carga de administración: Se define como el grado de trabajo y esfuerzo que deben realizar quienes administran un instrumento de medida (cuestionario), y depende del tiempo precisado para su administración, formación requerida del entrevistador y recursos necesarios para su aplicación y puntuación.

Carga de respuestas: Se entiende como el tiempo, esfuerzo y otros aspectos que se exigen a quienes se les administran un instrumento de medida (cuestionario).

CAT: Herramienta de lectura crítica consistente en un resumen breve de las pruebas científicas más relevantes en respuesta a una pregunta clínica concreta.

Causalidad: Indica que las asociaciones encontradas entre las exposiciones y los efectos son realmente de tipo causal y no consecuencia de sesgos ni del efecto de factores de confusión.

Cociente de probabilidad: Véase Razón de probabilidad.

Coefficiente de correlación: Índice que cuantifica la relación lineal entre un par de variables. Medida de asociación que indica el grado en el que dos variables continuas x e y poseen una relación lineal ($y = a + bx$). Su valor puede situarse entre -1 y $+1$. Cuando su valor es 0 no existe ningún grado de correlación. El coeficiente de este tipo que más se utiliza es el de Pearson

Coefficiente de Cronbach: Fórmula utilizada para la comprobación de la consistencia interna de un test que presupone que si varios ítems están midiendo el mismo atributo o dimensión, estos ítems deberían estar intercorrelacionados. Esta correlación la mediremos mediante el coeficiente alfa de Cronbach.

Coefficiente kappa: índice corregido por el azar de la concordancia entre, por ejemplo, juicios o diagnósticos que hacen 2 ó más observadores. El coeficiente adopta desde el valor 1 (cuando hay una concordancia perfecta) hasta el valor 0 cuando la concordancia observada es igual a la debida al azar.

Confusión: Situación en la que la estimación del resultado estudiado está sesgada por algún factor relacionado con la exposición y el resultado.

Consentimiento informado: Procedimiento formal para aplicar el principio de autonomía. Debe reunir por lo menos tres elementos: voluntariedad, información y comprensión. La voluntariedad implica que los sujetos puedan decidir libremente si quieren participar en un estudio, sin que haya persuasión, manipulación ni coerción. La información debe ser comprensible y debe incluir el objetivo del estudio y su procedimiento, los beneficios y riesgos potenciales, las posibles incomodidades derivadas de la participación y la posibilidad de retirarse del estudio en cualquier momento sin que ello de lugar a perjuicio, en este contexto es básica la comprensión de la información, que no sólo depende de esta, sino también de la competencia de quien la recibe, es decir de su capacidad para comprender.

Consistencia interna: Grado en que los ítems de una dimensión se correlacionan entre ellos mismos. Ver: Coeficiente de Cronbach.

Constructo: En psicometría, concepto abstracto no medible directamente que se quiere convertir en variable operativa medible. Generalmente, se utiliza para un concepto que existe en teoría, pero que no se puede observar directamente. Esencialmente, es una variable latente.

Correlación ítem-total:(Correlación total de las preguntas): Verifica la homogeneidad. Proceso mediante el cual comprobamos si cada ítem mide lo mismo que los demás y por lo tanto es sumable en una puntuación total que supuestamente mide el rasgo y que es la que después interpretamos y utilizamos.

Coste: Es el resultado del producto de la cantidad total del recurso consumido y el valor monetario de la unidad de dicho recurso.

Coste oportunidad: En economía de la salud, valor de las oportunidades perdidas por usar recursos en un servicio. Generalmente es cuantificado considerando el beneficio que se obtendría invirtiendo los mismos recursos en la mejor alternativa posible.

Criterios: Condiciones utilizadas como un estándar para juzgar los atributos.

Criterios de selección: Características sociodemográficas, clínicas, etc. que determina que los pacientes puedan ser o no incluidos en un estudio. Nos van a servir para definir la población de estudio.

Curva de Gauss: En matemáticas, la campana o curva de Gauss es la representación gráfica de la ecuación matemática que corresponde a una Distribución normal. Tiene forma de campana y debe su nombre al matemático alemán Carl Friedrich Gauss.

Curvas ROC: Representación gráfica completa de la capacidad de una prueba para discriminar lo normal de lo que no lo es.

Delimitador de campo: Ver Etiqueta.

DeCS - Descriptores en Ciencias de la Salud: Tesauro que puede considerarse, hasta cierto punto, como una versión en español del MeSH - Medical Subject Headings de la U.S. National Library of Medicine. Fue desarrollado a partir de él con el objetivo de permitir el uso de terminología común para búsqueda en tres idiomas (español, inglés y portugués), proporcionando un medio consistente y único para la recuperación de la información independientemente del idioma. Además de los términos médicos originales del MeSH contiene áreas específicas de Salud Pública y Homeopatía.

Descriptor: Término del lenguaje documental, carente de variabilidad, sinonimia y polisemia, escogido para representar el contenido de un documento, que puede ser utilizado en el análisis y recuperación de la información. Relacionado con Palabra clave.

Descriptor mayor: Ver Mayor Topic.

Descriptores en Ciencias de la Salud: Ver DeCS

Descuento: Procedimiento para calcular los costes o beneficios que ocurren en diferentes fechas a su valor en el momento presente, llamado valor actual.

Dimensiones: Aspectos relevantes que engloba el constructo o concepto que se mide a través del cuestionario.

Diseño: Conjunto de estrategias y procedimientos que se seguirán para dar respuestas al problema y comprobar la hipótesis. Engloba al selección del tipo de estudio, la población a estudiar, la definición de las variables y escalas de medición y los métodos de recogida, procesado y análisis de los datos.

Docencia virtual (“on-line” o “e-learning”): Definida como la utilización de las nuevas tecnologías multimedia y de internet para mejorar la calidad del aprendizaje, facilitando el acceso de recursos y servicios; así como los intercambios y la colaboración a distancia.

Dosis Diaria Definida (DDD): Dosis de mantenimiento promedio en la principal indicación de un fármaco en pacientes adultos. Se usa para cuantificar el consumo de medicamentos, sobre todo cuando se compara el consumo de diversos fármacos de un mismo grupo, o cuando se realizan comparaciones internacionales. No se ve afectada por diferencias de precios o de tamaño de las presentaciones. El consumo se suele expresar en número de DDD por 1.000 habitantes y por día, y esta medida da una idea del número promedio de personas tratadas diariamente con un fármaco o grupo de fármacos en una región o país.

Efectividad: Grado en que una determinada intervención origina un resultado beneficioso en las condiciones de la práctica clínica habitual, sobre una población determinada.

Efecto indeseado: Ver Reacción Adversa.

Eficacia: Grado en que una determinada intervención origina un resultado beneficioso en condiciones ideales de aplicación.

Eficiencia: Relación entre los efectos o resultados finales que se alcanzan con una intervención y coste que supone la misma (recursos humanos, materiales y de tiempo).

Encabezamientos de Materias Médicas: Ver MeSH.

Enmascaramiento: Procedimiento por el que se asegura que los sujetos participantes en un ensayo clínico, los observadores o ambos, no conocen el tratamiento u otra característica que pudiera sesgar los resultados. Se denomina simple ciego cuando el sujeto, pero no el observador ignora cuál de los posibles tratamientos recibe; doble ciego es cuando sujeto y observador ignoran el tratamiento recibido; y triple ciego es cuando el sujeto participante, el investigador observador y el investigador que analiza los datos ignoran el tratamiento recibido.

Ensayo clínico: “Toda investigación efectuada en seres humanos para determinar o confirmar los efectos clínicos, farmacológicos y/o demás efectos farmacodinámicos, y/o de detectar las reacciones adversas, y/o de estudiar la absorción, distribución, metabolismo y excreción de uno o varios medicamentos en investigación con el fin de determinar su seguridad y/o su eficacia” (Real Decreto de 16 de febrero de 2004).

Ensayo clínico controlado: Ensayo clínico en el que los sujetos son asignados a dos grupos: uno (grupo experimental) recibe el tratamiento que se está probando y el otro (grupo de comparación o control) recibe un tratamiento alternativo. En el ensayo clínico controlado aleatorizado la asignación a un tratamiento experimental o control se realiza al azar de forma que ni el sujeto ni el investigador responsable de su selección o tratamiento puedan influir en su asignación.

Entry Term: Relacionado con Palabra clave (véase).

Epidemiología: Estudio de la distribución y los determinantes de los estados o acontecimientos relacionados con la salud de poblaciones. Ciencia que investiga las causas y la historia de las enfermedades y el desarrollo de éstas en relación con determinadas intervenciones o exposiciones.

Epistemología: Doctrina de los fundamentos y métodos del método científico. La función de la estadística en el modelo empírico analítico.

Equivalencia: capacidad del instrumento de proporcionar los mismos valores en las mismas condiciones clínicas cuando las mediciones las hacen distintas personas.

Error aleatorio: Las diferencias de medidas entre la realidad y nuestra muestra y son debidas al azar.

Escala tipo Likert: La respuesta del individuo se da mediante una escala numérica, a la cual se le atribuyen las propiedades de las escalas de intervalo, permitiendo por tanto la realización de análisis estadísticos más complejos. Escalas que se utilizan con frecuencia en estudios de actitudes en los que las puntuaciones crudas se basan en respuestas alternativas graduadas a cada una de una serie de preguntas.

Especificidad: Referida a una prueba diagnóstica, es la proporción de personas verdaderamente no enfermas que son catalogadas como tales por medio de la prueba.

Estabilidad: Ver Reproducibilidad.

Estabilidad del test: Véase Método test-retest.

Estadística: Recopilación y organización de hechos o datos numéricos, sobre todo cuando hay muchos. Conjunto de métodos que ayudan a la toma de decisiones frente a la incertidumbre. La estadística no ofrece pruebas ni certidumbre, sino probabilidades.

Estadístico: Medida estadística (media, mediana, etc...) calculada sobre los datos de una muestra.

Estadístico kappa: Medida del acuerdo, corregida por el azar, aplicada a datos binarios u ordinales. Se calcula como la proporción de acuerdo, aparte del que ya sería de esperar por azar, que ha sido observado entre dos repeticiones del mismo instrumento.

Estudios ambispectivos: Estudios que se inician como los retrospectivos (Ver: Estudios retrospectivos) para el análisis de un fenómeno ya ocurrido y, posteriormente, continuamos con un seguimiento prospectivo hasta la aparición de un efecto a largo plazo (Ver: Estudios prospectivos).

Estudios analíticos: Estudios diseñados para examinar asociaciones, cuyo objeto final suele ser identificar o medir los efectos de factores de riesgo o de intervenciones específicas sobre la salud.

Estudios de caso-control: Estudios en los que la selección de la población a estudiar se hace en función de la presencia o ausencia de una enfermedad o efecto, eligiéndose un grupo de individuos que tienen la enfermedad o efecto (casos), y otro que no los presenta (controles).

Estudios de cohortes: Tipo de estudios en los que los sujetos del estudio se seleccionan a partir presencia o ausencia de la exposición que se trata de analizar y de la que se sospecha es responsable del efecto observado.

Estudio descriptivo: Estudio cuyo objetivo es constatar hechos, observar lo que ocurre en nuestra población. Permiten generar hipótesis.

Estudio experimental: Estudio en el que el investigador asigna y controla el factor de estudio.

Estudios de incidencia: Tipo de estudios en los que seguimos a una población determinada a lo largo del tiempo para analizar la aparición o no de un fenómeno de interés. La unidad de análisis es el individuo.

Estudios longitudinales: Estudios en los que la recogida de los datos de interés hace referencia a dos momentos separados en el tiempo (se analizan los sujetos en varios momentos a lo largo del tiempo).

Estudios observacionales: Estudios en los que el investigador no participa directamente en la intervención o exposición (factor de estudio), se limita a observar (registrar) lo que ocurre en la realidad.

Estudios pragmáticos: Estudios cuyo objetivo principal es la evaluación de la adecuación de la práctica asistencial en condiciones normales.

Estudios pre y post-intervención: Estudios en los que se registran la situación basal de la muestra y los cambios que experimenta tras su exposición a una determinada intervención.

Estudios de prevalencia: Estudios en los que se describe una serie de variables en una población determinada y en un momento determinado. La unidad de análisis es el individuo.

Estudios prospectivos: Tipo de estudios en los que los hechos a estudiar no han tenido lugar aun y los sujetos son seguidos hasta obtener el resultado.

Estudio retrospectivo: Estudio en el cual los hechos a estudiar ya han tenido lugar al iniciar el estudio.

Estudio transversal: Tipo de estudio en el que se estudia la exposición y el desenlace al mismo tiempo.

Etiqueta: Abreviatura del nombre de un campo que puede utilizarse en su lugar para dirigir la búsqueda de información. Sin: Tag, Delimitador de campo.

Evaluación crítica: Véase Lectura crítica.

Evaluación económica: Análisis comparativo de opciones alternativas en términos de sus costes y consecuencias. Los estudios de evaluación económica pueden ser parciales, cuando sólo se describen los costes, los resultados o ambos a la vez pero de una única opción o que comparan dos o más alternativas pero sólo en función de sus costes o de sus consecuencias sin valorar ambos a la vez, o completos, cuando se analizan comparativamente dos o más alternativas mediante la valoración de los costes y de las consecuencias de cada una de ellas.

Evaluación de tecnologías sanitarias: Proceso de análisis e investigación dirigido a estimar el valor y contribución relativos de cada tecnología sanitaria (véase) en la mejora de la salud individual y colectiva, teniendo en cuenta su impacto económico, social e, incluso, implicaciones éticas.

Explode: Obtención de las referencias que contienen el descriptor especificado o cualquiera de los descriptores de su mismo campo semántico pero de significado más restringido. Sin: Automatic explosion, Exploding, Explosion, Explosión, Explosión automática.

Exploding: Ver Explode.

Exploración complementaria: Conjunto de pruebas diagnosticas realizadas con la finalidad de confirmar una hipótesis diagnostica o hacer el seguimiento de la patologia de un paciente.

Explosión: Ver Explode.

Explosión automática: Ver Explode.

Factibilidad: Indica las posibilidades reales y práctica de que un instrumento de medida pueda ser contestado, completado y analizado. Está relacionad con el nivel de aceptabilidad del instrumento y se refleja en la proporción de rechazos, no respuesta y datos perdidos.



Factor de confusión: Variable que está simultánea e independientemente asociada al factor de riesgo estudiado y a la enfermedad estudiada, que puede modificar el resultado del estudio. Tal variable debe ser identificada y se debe evitar su influencia antes de la recogida de datos (mediante el emparejamiento o la restricción) o bien en la fase de análisis, mediante el análisis estratificado y la regresión múltiple.

Factor de impacto: Indicador bibliométrico, elaborado por el ISI (véase) y publicado en el Journal Citation Reports, que mide el uso de las revistas. Suele calcularse dividiendo el número de veces que son citados los artículos publicados en una revista durante un plazo de tiempo (habitualmente dos años), entre el número de artículos publicados por esa misma revista y en ese mismo plazo de tiempo. Puede modificarse para calcular el impacto de autores, instituciones... Sin: Impact Factor.

Factor de riesgo: Factor relacionado con el riesgo de que ocurra un suceso (p. ej. enfermar).

Falso negativo: Referido a una prueba diagnóstica, número de enfermos en que el resultado sería negativo.

Falso positivo: Referido a una prueba diagnóstica, número de sanos en que el resultado sería positivo.

Farmacovigilancia: Programas que proporcionan de forma continuada una evaluación sobre la seguridad de los fármacos. El factor clave en este sistema son los profesionales sanitarios que participan en el programa de notificación espontánea de reacciones adversas.

Fiabilidad: Grado de consistencia con que mide el atributo para el que está diseñado. La consistencia implica la ausencia de variabilidad o desacuerdo en repetidas observaciones o mediciones. Grado en que las mismas mediciones de los individuos que se obtienen en diferentes condiciones dan resultados similares.

Fiabilidad inter-observador (inter-ratio): Mide la variación que ocurre entre diferentes observadores al medir a un mismo individuo con un mismo instrumento .

Field: Ver Campo.

Filtros metodológicos: Son estrategias de búsqueda que nos ayudan a encontrar, en bases de datos como Medline, diferentes tipos de estudio sobre tratamiento, diagnosis, etiología o pronóstico. En Pubmed se denominan "Clinical Queries". Para aplicar uno de estos filtros sólo tenemos que combinar nuestra búsqueda con AND al filtro que nos interesa.

Forest/ Forest Plot: Tipo de gráfico ampliamente utilizado para representar los resultados de los diversos estudios seleccionados para una revisión sistemática. El Forest Plot ayuda a decidir si los resultados de los estudios individuales son suficientemente similares para poder ser combinados en una sola estimación.

Generabilidad: Véase Validez externa.

Gold Standard: Es el criterio que define quién tiene "realmente" la enfermedad (o condición) en estudio.

Grado o nivel de significación: En las pruebas de significación estadística, es el valor de p, el cual, en sentido estricto, en un ensayo clínico debe ser preespecificado en la fase de diseño, el nivel aceptado con mayor frecuencia es de 0.05. Ver: P (probabilidad), Significación clínica y Significación estadística.

Grupo de comparación: distintos grupos que intentamos comparar en un estudio analítico, ya sean estudios caso-control, estudios de cohortes o ensayos clínicos.

Grupo control: En un ensayo clínico, el grupo que recibe placebo o bien un tratamiento activo ya conocido, aceptado y establecido. Sin: Grupo de referencia.

Grupo intervención: En un ensayo clínico, el grupo que recibe el tratamiento que se desea evaluar.

Guía farmacoterapéutica: Documento que contiene una relación limitada de medicamentos recomendados para la prescripción en un ámbito determinado, seleccionada a partir de la oferta farmacéutica en función de unos criterios previamente establecidos, con la participación y el consenso de los profesionales a los que va destinada.

Guía de practica clínica: Conjunto de recomendaciones diseñadas para ayudar a los profesionales sanitarios y a los usuarios a seleccionar las opciones diagnósticas y/o terapéuticas más adecuada en el abordaje de una condición clínica específica.

Hipótesis: Suposición, fundada en observaciones o reflexiones, que puede conducir a predicciones refutables. También, conjetura elaborada de forma que pueda ser comprobada y refutada.

History: Archivo de todas las consultas bibliográficas realizadas a lo largo de una sesión en PubMed. También designa al subgrupo de registros que provienen de artículos sobre historia de la medicina.

Holística: Relativa al "holismo" que es la doctrina que propugna la concepción de cada realidad como un todo distinto de la suma de las partes que lo componen.

Impact Factor: Ver Factor de impacto.

Incremento absoluto del riesgo (IAR): Cuando el tratamiento experimental aumenta la probabilidad de un acontecimiento adverso, diferencia absoluta entre las tasas de acontecimientos adversos.

Incremento relativo del riesgo (IRR): Cuando el tratamiento experimental aumenta la probabilidad de un acontecimiento adverso, aumento proporcional de la tasas de acontecimientos adversos en el grupo experimental relativa a la tasa en el grupo de referencia.

Indicador de adhesión a la Guía Farmacoterapéutica: Indicador genérico de prescripción que indica el porcentaje de Dosis Diaria Definida (DDD, véase) de medicamentos prescritos incluidos en la Guía Farmacoterapéutica respecto del total de DDD de medicamentos prescritos.



Índice Bibliográfico Español en Ciencias de la Salud (IBECS): Base de Datos de la Biblioteca Nacional de Ciencias de la Salud (BNCS) del Instituto de Salud Carlos III (ISCIII), contiene referencias de artículos de revistas científico sanitarias editadas en España e incluye contenidos de las diferentes ramas de las ciencias de la salud tales como Medicina (incluyendo Salud Pública, Epidemiología y Administración Sanitaria), Farmacia, Veterinaria, Psicología, Odontología y Enfermería.

Índice Médico Español (IME): Base de datos referencial bibliográfica, creada y gestionada por el Instituto de Historia de la Ciencia y Documentación (IHCD), centro mixto de la Universidad de Valencia y del Consejo Superior de Investigaciones Científicas (CSIC). Cubre 321 publicaciones periódicas editadas en España especializadas en Biomedicina (Administración sanitaria, Farmacia clínica, Medicina experimental, Microbiología, Psiquiatría, Salud pública).

Indización: Proceso por el que el documentalista asigna unos descriptores que describen el contenido conceptual de documento con el fin de facilitar su recuperación

Index: Listado de todos los términos (unitérminos o compuestos) con significado que están incluidos en la Base de Datos. Esta opción permite hacer una búsqueda seleccionando los términos directamente del listado.

Indexar: Ver indización.

Ineficiencia: Situación en la cual los recursos no son empleados de la forma en que se incrementarían los beneficios producidos.

Inferencia: Proceso de pasar de las observaciones y axiomas a la generalizaciones. En estadística, desarrollo de la generalización a partir de los datos de la muestra, generalmente con un margen de incertidumbre.

Instrumentos de medida de CVRS específicos: Los instrumentos específicos incluyen dimensiones de la CVRS de una determinada afección que se quiere estudiar, capacidad funcional, o de un determinado grupo de pacientes.

Instrumentos de medida de CVRS genéricos: Los instrumentos de calidad de vida genéricos no están relacionados con ningún tipo de enfermedad o especialidad en particular. Son aplicables tanto a la población general como a grupos específicos de pacientes.

Interpretabilidad: grado en el que uno puede asignar un significado comprensible a las puntuaciones cuantitativas del instrumento.

Intervalo de confianza: Margen o recorrido comprendidos entre dos valores extremos (el límite inferior y el superior), entre los que cabe esperar el valor medio real de la población con un margen grado prefijado de seguridad. Ver: Nivel de confianza.

Intervención: Cualquier acción (preventiva, diagnóstica, sintomática o curativa) aplicada con el fin de modificar el curso de la situación previa. Cualquier tratamiento o actuación sanitaria que queramos evaluar

ISI: El Institute for Scientific Information fue creado por Eugene Garfield en 1958 con el objeto de publicar el primer Current Contents que cubría unas 200 revistas. En 1971, cubría más de 4000 revistas, publicaba más de 650 páginas de contenidos a la semana, y era consultado por más de un millón de científicos. En 2001, se desarrolló el ISI Web of KnowledgeSM, que integra literatura científica, tanto revistas (journals) como actas (proceedings), patentes, compuestos químicos, y hasta secuencias genómicas. El ISI fue adquirido por The Thomson Corporation, que lo convirtió en el Thomson Institute for Scientific Information. Es, por tanto, una institución privada que produce diversas Bases de Datos de primer orden en calidad y utilidad en la investigación científica, pero que sobre todo ha obtenido un notable éxito por la difusión de lo que se conoce como Impact Factor o Factor de Impacto.

ISSN: International Standard Serial Number, número internación único que identifica de manera inequívoca cada revista.

Ítem: Cada uno de los elementos, preguntas o cuestiones que constituyen un test.

Key word: Ver Palabra clave.

Lectura crítica: Es el proceso de evaluar e interpretar la evidencia aportada por la literatura científica, considerando sistemáticamente los resultados que se presenten, su validez y su relevancia para el trabajo propio. Sin: Evaluación crítica.

Lenguaje documental: Lenguaje artificial creado específicamente para el análisis y recuperación de la información. Contrapuesto al Lenguaje natural, se diferencia de este especialmente por carecer de variabilidad, sinonimia y polisemia.

Lenguaje natural: Expresión de ideas o de conceptos con los términos y con las reglas propios de un idioma o de un grupo especial de hablantes (p.e. sanitarios). Se contrapone a Lenguaje artificial y, específicamente en documentación, a Lenguaje documental.

Límites: Opción que ofrecen varias bases de datos, como Medline, que permite limitar/restringir una búsqueda de acuerdo a unas características como p.ej. artículos publicados en un idioma determinado, tipo de publicación (revisión, ensayo clínico), o grupo de edad, sexo, etc.

Likelihood Ratio: Véase Razón de probabilidad.

Literatura inédita: Literatura que no ha sido publicada.

Mapeo Automático de Términos: Ver Acotado Automático de Términos.

Mayor Topic: Materia principal de un artículo. Sin: Descriptor mayor.

Médical Subject Headings: Ver MeSH

Medicina Basada en la Evidencia (MBE): Práctica de la medicina basada en la integración de la maestría clínica de cada médico (es decir, el dominio del conocimiento y el juicio adquirido por la experiencia) con las mejores pruebas clínica externas disponibles. También ha sido definida como la utilización consciente y juiciosa de las mejores demostraciones provenientes de la investigación clínica para la toma de decisiones en el cuidado de cada paciente.

Medicamentos genéricos (Especialidad Farmacéutica Genérica, EFG): Especialidad con la misma forma farmacéutica e igual composición cualitativa y cuantitativa en sustancias medicinales que otra especialidad de referencia, cuyo perfil de eficacia y seguridad está suficientemente establecido por su continuado uso clínico. La EFG debe demostrar la equivalencia terapéutica con la especialidad de referencia mediante los correspondientes estudios de bioequivalencia.

Medicamentos huérfanos: Cuando un fármaco no es desarrollado porque el detentador de la patente considera que no recuperará los costes de este desarrollo (en el caso de enfermedades raras y otras enfermedades no raras pero que ocurren en países pobres en los que no existe un “mercado” potencial), se lo denomina medicamento huérfano.

MeSH (Médical Subject Headings, Encabezamientos de Materias Médicas): Tesauro (diccionario que permite conocer la equivalencia entre los términos del lenguaje natural y los del lenguaje documental) de Medline y de PubMed.

MeSH Subheading: Subencabezamiento propio de MeSH.

MeSH Term: Descriptor propio de MeSH.

Meta-análisis: Integración estructurada y sistemática de la información obtenida en diferentes estudios sobre un problema determinado. Consiste en identificar y revisar los estudios controlados sobre un determinado problema, con el fin de dar una estimación cuantitativa sintética de todos los estudios disponibles.

Método matricial multirasgo-multimetodo (multitrait-multimethod matriz method, MTMM): procedimiento empleado para la validación de constructo aplicando para ello los conceptos de convergencia (validez convergente) y discriminabilidad (validez discriminante).

Método test-retest: Método utilizado para la comprobación de la fiabilidad de un test. Correlación entre la puntuación de dos administraciones de una prueba a los mismos individuos. De esta manera se mide uno de los conceptos incluidos en la fiabilidad: la estabilidad del test.

Modelo de medida: Se refiere a las escalas y subescalas de un cuestionario y a los procedimientos seguidos para crear las puntuaciones.

Muestra: Fracción seleccionada de una población determinada, idealmente representativa de la población de referencia y de tamaño suficiente.

Muestreo: Selección de los sujetos de una población que tomarán parte en un estudio. El muestreo aleatorio implica que cada participante potencial de la población estudiada tiene las mismas probabilidades de ser incluido en la muestra.

NEAR: Operador booleano utilizado en las búsquedas en Librería Cochrane Plus. Buscará documentos donde las dos palabras que enlaza NEAR estén separadas entre sí por un máximo de seis palabras en cualquier orden.

NEXT: Operador booleano utilizado en las búsquedas en Librería Cochrane Plus. Recupera documentos donde las dos palabras están juntas en ese orden.

Nivel de confianza: Se expresa en términos probabilísticos y se interpreta como la probabilidad de que el verdadero parámetro poblacional esté dentro del intervalo de confianza. Ver: Intervalo de confianza.

Normograma de Fagan: Método de cálculo de la probabilidad posttest mediante la representación gráfica de una tabla. Permite conocer la probabilidad posttest mediante el trazo de una línea que una la probabilidad pretest y la razón de probabilidad, el siguiente punto de intersección con la línea situada a la derecha, señala la probabilidad posttest.

Número necesario para dañar (number needed to harm, NNH): Número de pacientes que deben tomar el tratamiento experimental para que se produzca un caso adicional de un determinado acontecimiento adverso, en comparación con los pacientes que reciben el tratamiento de referencia.

Número necesario para tratar (NNT): Es una medida de la eficacia de un tratamiento. Cuando el tratamiento experimental incrementa la probabilidad de un acontecimiento favorable (o cuando disminuye la de un acontecimiento adverso), número de pacientes que hay que tratar para dar lugar a un paciente más con mejoría (o para prevenir un acontecimiento adverso adicional). Se calcula como $1/\text{RAR}$.

Odds: Se llama odds a la razón en la que el numerador es la probabilidad de que ocurra un suceso y el denominador es la probabilidad de que tal suceso no ocurra.

Odds post-test: Equivale a los odds (proporción) de que un paciente tenga la enfermedad una vez que se ha realizado el test diagnóstico.

Odds pre-test: Definida como los odds (proporción) de que un paciente tenga la enfermedad antes de que el test diagnóstico se efectúe.

Odds ratio (OR): Resultado de dividir la odds del grupo de interés (expuesto) por la odds del grupo de referencia (control). Si el $\text{OR} = 1$ el riesgo o probabilidad de que ocurra el evento es el mismo para ambos grupos (el efecto del tratamiento experimental es el mismo que el del tratamiento control); si el $\text{OR} > 1$ se interpretará como que el tratamiento experimental favorece la aparición del evento y si el $\text{OR} < 1$ diremos que protege de la aparición del mismo.

Operadores booleanos: Partículas habitualmente expresadas en inglés (AND, OR, NOT) que indican operaciones lógicas entre conceptos. Las operaciones son las de intersección (AND = recupera los documentos que contengan ambos términos), adición (OR = recupera los documentos que contengan tanto un término como otro, o la suma de los dos) o exclusión (NOT = sirve para excluir documentos que tienen un determinado término).

P: Seguía de la abreviatura n.s. (no significativa) o del símbolo $<$ o una cifra decimal, indica la probabilidad de que la diferencia observada en una muestra haya ocurrido puramente por azar siendo los grupos comparados realmente semejantes, es decir bajo la hipótesis nula.



Palabra clave: Término propio del lenguaje natural, sin otras normas que las del habla de la cual procede, que puede utilizarse en la búsqueda de información. Con frecuencia se utiliza como sinónimo de Descriptor, pero este uso no es correcto y conviene evitarlo. En MeSH las palabras clave son el equivalente a los términos de entrada (Entry Terms) de un diccionario de idiomas. Sin: Key word.

Palabra de texto: PubMed considera “palabras de texto” a todos los números y términos con significado que aparecen, entre otros menos importantes, en los campos título (title [TI]), resumen (abstract [AB]), descriptores (MeSH terms [MH]), subencabezamientos (MeSH Subheadings [SH]), así como los nombres de sustancias químicas (Substance Name [NM]) y los nombres de personas (Personal Name [PS]) cuando unas y otras son la materia del artículo. Sin: Text Word.

Panel de expertos: Reunión de un grupo de profesionales de prestigio en una determinada materia que abordan, exponen, contrastan ... sus opiniones y experiencias sobre la actuación en un tema preocupante de interés para el que se busca la mayor uniformidad de criterios que no es posible obtener por otro tipo de estudios más eficientes.

Perdidas de seguimiento: Aquellos pacientes que son perdidos durante el estudio y en los cuales no podemos valorar los resultados finales.

Población: Grupo individuos, objetos, mediciones, etc., con determinadas características demográficas, de donde se obtiene la muestra y a la que se quiere hacer inferencia de los resultados de un estudio.

Positivismo lógico: También denominado neopositivismo. Movimiento filosófico contemporáneo que destaca en filosofía la importancia del análisis del lenguaje y de la metodología científica.

Pretest cognitivo: Conjunto de estrategias y procedimientos que se utilizan para determinar si el cuestionario funciona tal y como lo ha concebido el investigador. Debe reproducir exactamente las condiciones en las que se responderá el cuestionario definitivo y, por tanto, tiene que realizarse con el cuestionario terminado y empleando el mismo procedimiento de recogida de información, puesto que la única diferencia entre el pretest y el trabajo de campo es el número de cuestionarios contestados.

Prevalencia: Número de casos de una enfermedad en una población y en un momento dados.

Prevalencia del test: Número total de positivos obtenidos al realizar una prueba diagnóstica en relación al total de sujetos que componen la muestra estudiada.

Probabilidad: Es el cociente entre los casos posibles de que ocurra un hecho dividido por el total de casos.

Probabilidad pretest o preprueba: Proporción de pacientes que tienen la enfermedad en la muestra estudiada, y cuyo diagnóstico es fehacientemente establecido por un estándar de referencia. Es la prevalencia de la enfermedad.

Probabilidad postest o posprueba: Probabilidad de tener una enfermedad tras la realización de una prueba diagnóstica.

PubMed: Base de datos de bibliografía sobre ciencias de la salud. Puede obtenerse de ella información acerca de publicaciones de carácter científico y de contenido médico (en su sentido más amplio, incluyendo enfermería, odontología, veterinaria...).

Randomización: Ver Aleatorización.

Razón de probabilidad: Probabilidad de un resultado en los enfermos dividido por la probabilidad de ese mismo resultado en no enfermos. Sin: Cociente de probabilidad, Likelihood Ratio.

Reacciones adversas a medicamentos (RAM): Según la OMS, “reacción nociva y no deseada que se presenta tras la administración de un fármaco, a dosis utilizadas habitualmente en la especie humana, para prevenir, diagnosticar o tratar una enfermedad, o para modificar cualquier función biológica”. Las reacciones adversas se clasifican en dos grandes grupos: RAM tipo A, de tipo cuantitativo o predecibles, y RAM tipo B, de tipo cualitativo o no predecibles. Sin: Efecto indeseado.

Record: Ver Asiento.

Reducción absoluta del riesgo (RAR): Diferencia aritmética absoluta entre la tasa de acontecimientos (riesgo) en el grupo experimental y la tasa de acontecimientos (riesgo) en el grupo control. También se puede denominar: Diferencia de riesgos.

Reducción relativa del riesgo (RRR): Cociente entre la reducción absoluta del riesgo (RAR) y el riesgo en el grupo control. A la reducción relativa del riesgo se le llama también “fracción atribuible” o “diferencia relativa del riesgo”.

Registro: Ver Asiento.

Regresión logística: Tipo de análisis de regresión (véase) en el que la variable dependiente sólo puede tomar dos valores. En el análisis de regresión la variable dependiente puede ser o bien categórica o bien continua (véase).

Repetibilidad: Referida a una prueba, es el grado en que los resultados son idénticos o estrechamente semejantes cada vez que aquella se lleva a cabo por los mismos profesionales y los mismos emplazamientos.

Reproducibilidad: Grado en que los resultados son idénticos o estrechamente semejantes cada vez que una prueba se lleva a cabo por distintos profesionales y es distintos ámbitos. Sin: Estabilidad.

Retroalimentación: Método para mantener la acción o eficacia de un sistema mediante la continua revisión de los elementos del proceso y de sus resultados con el fin de realizar las modificaciones necesarias.

Reunión educacional: Visita individualizada de un educador a un profesional de la salud en su ámbito de trabajo; en ella se suelen combinar técnicas de auditoria y retroalimentación.

Revisión: Se denomina así a todo intento que sintetice los resultados y las conclusiones de dos o más publicaciones relacionadas con un tema dado.

Revisión bibliográfica: recopilación de literatura realizada sobre un tema concreto en un intervalo de tiempo concreto del que se realiza un resumen y en el que no quedan explícitos los criterios de búsqueda y se destacan las conclusiones a criterio del autor.

Revisiónes narrativas: Se basan en el análisis que hace un experto de los estudios originales existentes sobre un determinado tema o problema para posteriormente subrayar aquellos resultados que considera más importantes.

Revisiónes sistemáticas: Revisión en las que la evidencia sobre un tema ha sido sistemáticamente identificada, criticada y resumida de acuerdo a unos criterios predeterminados.

Riesgo: Resultado predecible o incierto desfavorable de una actividad, intervención o exposición. Probabilidad de que ocurra un determinado fenómeno adverso.

Riesgo relativo (RR): Cociente entre el riesgo de padecer un acontecimiento entre los individuos expuestos a un factor de riesgo determinado y el riesgo de padecerlo entre los no expuestos. También se puede denominar: Cociente de riesgos.

Seguimiento: Periodo de duración de un estudio durante el cual se mantienen contactos periódicos con los participantes en el mismo con el fin de evaluar las variables de estudio.

Sensibilidad: Referida a una prueba diagnóstica, es la proporción de personas verdaderamente enfermas que ha sido catalogadas como tales mediante dicha prueba.

Sensibilidad al cambio (de un instrumento): se refiere a la medida en que detecta y mide de manera fiable las variaciones más pequeñas de un atributo, rasgo, dimensión o constructo.

Serie de casos: Tipo de estudio que describe una serie de características poco conocidas de un proceso o tratamiento y facilitan el conocimiento del mismo y su evolución en el tiempo.

Sesgo: Error sistemático producido en la estimación de los resultados, a causa de defectos en el diseño del estudio, recogida de datos o seguimiento.

Sesgo de desgaste: Diferencias sistemáticas que se producen entre los grupos de un ensayo clínico por la pérdida de participantes durante el estudio.

Sesgo de información: Sesgo relacionado con la forma en la que se obtuvieron los datos. Entre ellos se incluyen los errores cometidos en la valoración del estado de los pacientes al inicio, durante o al final del estudio; también los que ocurren cuando los datos se obtienen de fuentes sin que se conozca la calidad de las mismas. Algunos de los sesgos de información serían: el sesgo del recuerdo o de memoria, sesgo de sospecha diagnóstica, sesgo del entrevistador...

Sesgo de interpretación de las pruebas: Error sistemático que surge debido a la realización de las pruebas de forma no independiente y a ciegas con la consecuencia de que el conocer los resultados de una prueba influye en los de la otra.

Sesgo de publicación: Sesgo que surge debido a la tendencia a publicar sólo estudios con resultados positivos.

Sesgo de selección: Error sistemático consecuencia de que las características de los sujetos incluidos en un estudio son diferentes de las características de los no incluidos, de modo que la muestra no es representativa de la población de referencia. Se incluyen en este grupo el sesgo de los voluntarios, el de prevalencia-incidencia, el de los emigrantes, el de selección no aleatorizada, el de Berkson, el de pertenencia a grupo, el del trabajador sano...

Sesgo de sospecha diagnóstica: Sesgo que ocurre cuando la interpretación de una prueba se haga conociendo el resultado de otra.

Sesgo de reproducibilidad: Sesgo debido a la ausencia de información detallada de las condiciones en que se ha aplicado la prueba, así como las definiciones de los resultados positivos y negativos.

Sesgo de resultados no interpretables: Error sistemático consecuencia de no incluir pacientes en diferentes estadios de la enfermedad.

Sesgo de verificación o de confirmación diagnóstica: Sesgo que ocurre cuando el estudio se limita a individuos sobre los que se realizó con anterioridad la prueba "gold Standard", que suelen ser los de mayor probabilidad de padecer enfermedad.

Sesgo en el espectro de la enfermedad: ¿

Sesgo por ausencia de "gold Standard": Error sistemático que se produce cuando entre los criterios de referencia que se emplean en la comparación de la utilidad de una prueba diagnóstica (en ausencia de "gold Standard") se incluye la prueba de estudio.

Sesión bibliográfica: Herramienta docente consistente en reuniones entre profesionales que realizan una lectura crítica de un artículo científico.

Significación clínica: Probabilidad de que una diferencia observada tenga una repercusión sobre el curso del problema o enfermedad tratados que sea relevante para un paciente dado o para un conjunto de pacientes.

Significación estadística: Probabilidad de que una diferencia observada sea resultado de la casualidad y no de los determinantes causales en un estudio. El hallazgo de una significación estadística no implica necesariamente significación clínica.

Síntesis: Fuentes de información que recogen resúmenes estructurados de revisiones sistemáticas de alta calidad o de artículos originales.

Síntesis: Fuentes de información que incluyen revisiones sistemáticas de alta calidad metodológica.



Sistema: Fuente de información que se caracteriza por abarcar una gran variedad de temas, proporcionar un resumen de los resultados de revisiones de alta calidad y ser actualizado con frecuencia.

Subencabezamiento: Término que delimita el significado del descriptor al que acompaña. P.e. “Drug therapy” (tratamiento farmacológico) como subencabezamiento unido al descriptor Hypertension nos informaría que un artículo trata, entre todos los aspectos posibles de la hipertensión, su tratamiento farmacológico. Sin: Subheading.

Subheading: Ver subencabezamiento.

Tag: Ver Etiqueta.

Tamaño de la muestra: Número mínimo necesario de sujetos que participan en un estudio, pero a su vez suficiente para proporcionar una respuesta fiable al objetivo final de dicho estudio. La determinación de dicho tamaño mínimo es de naturaleza estadística y depende de varios factores.

Técnica de grupos conocidos: Procedimiento utilizado para la validación de constructo según la cual el instrumento se aplica a los grupos que, según se presume, diferirán respecto del atributo crítico a causa de una característica conocida.

Tecnología sanitaria: Conjunto de medicamentos, aparatos, procedimientos médicos y quirúrgicos usados en la atención médicos y los sistemas organizativos y de soporte con los que se presta la atención sanitaria.

Teoría estadística de Bayes o Teorema de Bayes: Establecimiento inductiva de la probabilidad de una hipótesis a partir de algunas observaciones. Según el teorema de Bayes, la probabilidad posterior (o a posteriori) de una hipótesis es igual a una constante, multiplicada por la verosimilitud o posibilidad de la hipótesis (definida como la probabilidad de observar unos acontecimientos determinados si la hipótesis considerada fuera de hecho cierta) y por la probabilidad anterior de la hipótesis.

Término MeSH: Ver MeSH Term

Término de entrada: Relacionado con Palabra clave (véase).

Tesaurus: Es un repertorio de términos con expresión de las relaciones semánticas que los unen. Habitualmente consta de dos partes, una sección categorizada, que expresa las relaciones jerárquicas, genérico-específicas, entre los términos normalizados (descriptores), lo que nos permite conocer el campo semántico (significado) que abarcan los distintos términos. La sección alfabética es una ordenación de todos los términos, normalizados (descriptores) o no normalizados (palabras clave), que de forma similar a un diccionario de idiomas, permite conocer las relaciones de equivalencia entre unos y otros términos. Sin: Thesaurus.

Test Chi cuadrado (2): Prueba estadística utilizada para determinar el grado de asociación entre variables categóricas (véase).

Test de McNemar: Prueba de significación estadística para probar la hipótesis nula de inexistencia de cambios en la proporción de sujetos que experimentan un acontecimiento, cuando cada individuo es evaluado dos veces (en condiciones diferentes) y los datos están emparejados.

Test T de Student: Prueba de significación estadística paramétrica (véase) para contrastar la hipótesis nula respecto a la diferencia entre dos medias.

Test U de Mann-Whitney: Prueba de significación estadística no paramétrica (véase) para probar la hipótesis nula de que el parámetro de localización (generalmente la mediana) es el mismo cuando se comparan dos grupos independientes, cualquiera que sea el tipo de distribución de la variable (distribución normal o de otro tipo).

Test de Wilcoxon: El contraste de Wilcoxon es la técnica no paramétrica paralela al test T de Student para muestras apareadas.

Text Word: Ver Palabra de texto.

Thesaurus: Ver Tesoro.

Transferabilidad: Término acuñado en investigación cualitativa sugiere que en la medida en que el investigador ha respetado el criterio de maximizar la diversidad en la obtención de la muestra y realice una detallada descripción del contexto y participantes, los hallazgos pueden ser aplicables en contextos similares.

Triangulación: Técnica empleada en investigación cualitativa para controlar sesgos. Consiste en obtener distintas perspectivas del fenómeno estudiado utilizando diferentes métodos de investigación (cualitativos y cuantitativos). La triangulación puede ser simultánea o secuencial.

Truncado: El término truncar significa en español "cortar el extremo de una cosa". Aplicado a un término de búsqueda consiste en separar del término su terminación y sustituirla por un símbolo comodín (en PubMed es un asterisco *-). El resultado es la selección de todos los términos que contienen la raíz superviviente.

Umbral de decisión: Nivel de probabilidad a partir del cual el beneficio previsible de toma una decisión de intervenir iguala o mejora al beneficio de no tomarla

Validez: Capacidad de un instrumento para medir aquello para lo que ha sido diseñado, denota el grado en que el instrumento mide lo que se supone que debe medir.

Validez aparente o facial: Forma especial de la validez de contenido que se refiere al hecho de que el instrumento parezca, de la impresión a las personas evaluadas, de que es adecuado para medir lo que se pretende medir.

Validez concurrente: Denota la capacidad de un instrumento de distinguir a los individuos que difieren en su condición actual respecto de un criterio.

Validez convergente: Aspecto de la validez de constructo que se refiere al grado en que diferentes formas de medir el mismo constructo (o dimensiones de un constructo) se correlacionan la una con la otra.

Validez de constructo: Grado en que una prueba o instrumento mide un determinado rasgo, característica o constructo.

Validez de contenido: Determina si el instrumento contiene ítems representativos de todas las dimensiones que forman la definición del concepto o constructo en estudio.

Validez de criterio: Concordancia entre los resultados de la medida del instrumento y los resultados de la medida realizada mediante un criterio o patrón de referencia (gold standard) de lo que se quiere medir.

Validez discriminante o divergente: Aspecto de la validez de constructo que demuestra que una medida no se correlaciona con otras medidas cuya finalidad es medir rasgos, dimensiones o constructos diferentes a los que ella mide.

Validez externa: Grado en que las conclusiones obtenidas con la muestra de población que participa en un estudio pueden ser generalizadas a su población de referencia o a otras poblaciones, lugares, momentos e investigaciones.

Validez interna: Grado en que los resultados de un estudio miden la realidad para la muestra de población que participa en un estudio. Grado en que un estudio está libre de sesgos.

Validez predictiva: Se refiere a la idoneidad de un instrumento para diferenciar entre el desempeño o las conductas de los individuos respecto de un criterio futuro.

Valor global de un test: Proporción de sujetos correctamente clasificados por el test en relación al total de sujetos que componen la muestra estudiada.

Valor predictivo negativo: Referido a una prueba diagnóstica, la probabilidad de que una persona con un resultado negativo sea un verdadero negativo.

Valor predictivo positivo: Referido a una prueba diagnóstica, la probabilidad de que una persona con un resultado positivo sea un verdadero positivo.

Variabilidad: Variación entre individuos debido a determinantes genéticos, ambientales, patológicos...(variabilidad interindividual) u oscilación de variables biológicas en un mismo individuo según las circunstancias (variabilidad intraindividual).

Variable: Cualquier atributo, fenómeno o hecho que pueda tener diferentes valores.

Variable categórica: Se dice de la variable que presenta interrupciones. Una variable es categórica si, entre varios valores potencialmente observables, hay un valor que no puede ser observado. Ejemplo: variables binarias.

Variable continua: Se dice de la variable que no presenta interrupciones. Una variable es continua si, entre dos valores observables, siempre existe la posibilidad de que haya otro valor observable. Ejemplos: edad, peso, altura.

Verbatim: Su significado es: Literalmente, o palabra por palabra. Anglicismo utilizado en investigación cualitativa para denominar a la transcripción exacta, en la fase de exposición de resultados, de las expresiones surgidas durante el grupo de discusión de los participantes, es decir conservado la expresión literal grabada.

Verdaderos negativos: Grupo de pacientes sanos en los que el test es negativo.

Verdaderos positivos: Grupo de pacientes con enfermedad y que además son detectados por el test.

