



COMPRIMIDOS VAGINALES

misofar[®] 200

MISOPROSTOL 200mcg

Misofar[®] es un **análogo sintético de la PGE1** indicado para la **dilatación del cérvix de útero no grávido** previa a una histeroscopia u otros procedimientos ginecológicos que requieran acceder a la cavidad uterina¹

La aplicación vaginal de MISOFAR[®] 200:

- Disminuye** la resistencia cervical^{2,3}
- Reduce** la necesidad de utilizar dilatación cervical²
- Facilita** el proceso de introducción del histeroscopio⁴
- Disminuye** el dolor durante el proceso de histeroscopia³

pauta posológica

DOSIS RECOMENDADA: 400 mcg de misoprostol



En úteros **CON** historial de cesárea previa o cicatrices uterinas¹
2 a 4 horas antes

En úteros **SIN** historial de cesárea previa o cicatrices uterinas¹
2 a 8 horas antes

PRODUCTO HOSPITALARIO

Presentación en envase de **4 comprimidos vaginales¹**

- Sin necesidad de conservación en frío¹
- Comprimidos termoestables
 - Preservados en blíster de aluminio



Investigar y publicar en ginecología

Metodología y estadística

J. González de Dios • M.^a González-Muñoz



2E15MPEMI05 06/15

1. Ficha técnica y prospecto. 2. Barcaite E, et al. Vaginal misoprostol for cervical priming before hysteroscopy in perimenopausal and postmenopausal women. International Journal of Gynecology and Obstetrics (2005). 3. Guy Waddell, MD et al. Cervical Ripening Using Vaginal Misoprostol before Hysteroscopy: A Double-blind Randomized Trial. Journal of Minimally Invasive Gynecology (2008) 15, 739-744. 4. Preuthippan S, et al. Vaginal misoprostol for cervical priming before operative hysteroscopy: a randomized controlled trial. Obstet Gynecol. 2000; 96: 890-4.

Investigar y publicar en ginecología

Metodología y estadística

Javier González de Dios, María González-Muñoz*

Servicio de Pediatría. Hospital General Universitario de Alicante. Departamento de Pediatría.
Universidad Miguel Hernández. Alicante.

*Facultad de Medicina y Odontología.

Universidad de Valencia

Sumario

Claves estadísticas en la elaboración de artículos científicos:
algo más que números

2

Medidas comunes en epidemiología:
medidas de riesgo y de impacto

8

Tipos de estudios epidemiológicos:
entre la causalidad y la «evidencia»

11

Errores en los estudios epidemiológicos:
evitar el error aleatorio y el error sistemático

16

Las listas guía de comprobación:
un mar de siglas para navegar mejor en metodología
y estadística

19

Claves estadísticas en la elaboración de artículos científicos: algo más que números

La estadística es la parte más ardua de un artículo científico y la que más nos cuesta: nos cuesta enfrentarnos a ella como autores, tanto como entenderla como lectores. Sin embargo, es un apartado esencial, que fundamenta el rigor científico del estudio y, por ende, la validez de las conclusiones.

En el último medio siglo, el uso de la estadística se ha promovido y cuidado mucho en las principales revistas científicas. Pero una consecuencia desafortunada de este fenómeno ha sido que se ha puesto excesivo énfasis en examinar las pruebas de hipótesis con un enfoque excesivamente «estadístico», menos «clínico»: los datos se examinan en relación con una hipótesis nula, práctica que ha llevado a la creencia errónea de que el objetivo de los estudios debe ser obtener la diferencia «estadísticamente significativa», cuando en realidad el objetivo real debe ser determinar la magnitud de algún factor de interés y la diferencia «clínicamente importante».

Conocer el tema de un artículo científico ayuda a entender el trabajo, y conocer la estadística ayuda a entender la metodología. En realidad, la capacitación formal en bioestadística de los investigadores médicos no siempre ha sido la deseable; por ello, el hecho de que un artículo haya sido publicado en una revista biomédica, previa revisión por correctores, no significa que el artículo no tenga errores en la metodología estadística, salvo honrosas excepciones (hay revistas de prestigio que disponen de un consejo de revisores especializados en metodología y estadística).

Por todo ello, conviene recordar algunas premisas en relación con los aspectos estadísticos que deben tenerse en cuenta en la elaboración de un artículo científico, premisas que aquí expondremos como un decálogo práctico (tabla 1)¹⁻³:

1. Describir los métodos científicos con suficiente detalle. El objetivo es que un lector que tenga acceso a los datos originales pueda verificar los resultados notificados. Los estudios estadísticos se clasifican en diversas categorías según los criterios de Emerson y Colditz⁴ (tabla 2), autores que introdujeron un término útil, el de «accesibilidad estadística», entendido como un índice de complejidad de las técnicas estadísticas en los trabajos científicos. Y, así, se

Tabla 1. Decálogo para sobrevivir a la estadística

1. Describir los métodos científicos con suficiente detalle
2. Cuantificar los resultados
3. Conocer el significado del valor de p
4. No depender exclusivamente del valor de p
5. Proporcionar los detalles del proceso de aleatorización
6. Proporcionar los detalles del proceso de enmascaramiento (cegamiento)
7. Especificar el número de observaciones e indicar las pérdidas
8. Especificar el programa estadístico y las referencias sobre estadística
9. Definir los términos y evitar el uso no técnico de términos estadísticos
10. Controlar los errores de la investigación clínica

Tabla 2. Categoría de análisis estadístico*

Categoría	Descripción	
0	Ningún estudio estadístico	
1	Sólo estadística descriptiva	Por ejemplo, porcentajes, medias, desviaciones típicas, errores estándares, etc.
2	Prueba de la t de Student y pruebas Z	Para una o dos muestras (datos apareados y/o independientes)
3	Tablas bivariadas	Por ejemplo, prueba de la χ^2 , prueba exacta de Fisher, prueba de McNemar
4	Pruebas no paramétricas	Por ejemplo, prueba U de Mann-Whitney, prueba T de Wilcoxon
5	Estadísticos demoepidemiológicos	Por ejemplo, riesgo relativo, <i>odds ratio</i> , sensibilidad, especificidad
6	Correlación lineal de Pearson	Correlación clásica producto-momento (r)
7	Regresión simple	Regresión de mínimos cuadrados con una variable predictora y una variable de respuesta
8	Análisis de la variancia	Análisis de la variancia y de la covariancia: pruebas F
9	Transformación de variables	Uso de transformaciones (p. ej., logarítmicas)
10	Correlación no paramétrica	Por ejemplo, prueba rho de Spearman, tau de Kendall, pruebas de tendencia
11	Regresión múltiple	Por ejemplo, regresión polinómica y regresión paso a paso
12	Comparaciones múltiples	Por ejemplo, pruebas de Bonferroni, Scheffé, Duncan, Newman-Keuls
13	Ajustes y estandarización	Estandarización de tasas de incidencia y prevalencia
14	Tablas multivariadas	Por ejemplo, procedimiento de Mantel-Haenszel, modelos log-lineales
15	Potencia y tamaño muestral	Determinación del tamaño muestral a partir de una diferencia detectable (o útil)
16	Análisis de la supervivencia	Por ejemplo, tablas de vida (Kaplan-Meier), regresión de supervivencia (logística, de Cox) y otros (Breslow, Kruskal-Wallis, <i>log-rank test</i>)
17	Análisis económicos	Por ejemplo, coste-efectividad, coste-beneficio, coste-utilidad
18	Otros análisis diversos	Por ejemplo, análisis de sensibilidad, análisis discriminante, análisis <i>cluster</i> u otros modelos matemáticos

*Modificada de Emerson y Colditz⁴.

considera que el umbral básico para un lector estándar (aquel que ha superado un curso básico de bioestadística durante su currículo biomédico) se situaría en la estadística bivariada (hasta la regresión simple).

Si la prueba estadística es de uso común (p. ej., t de Student, χ^2 , Wilcoxon, Mann-Whitney, etc.), sólo cabe citarlo, pero en pruebas estadísticas de mayor complejidad (p. ej., análisis de la variancia, estudio multivariante, etc.) sí deben especificarse de forma clara. Conviene señalar si las observaciones son no pareadas (p. ej., comparación de dos medias entre dos grupos de pacientes que reciben diferentes tratamientos), o bien pareadas (p. ej., comparaciones de dos medias en el mismo grupo de pacientes en dos circunstancias diferentes, antes y después del tratamiento), porque variará el tipo de prueba estadística. Así, la prueba de Mann-Whitney utilizada para datos no pareados se sustituye por la prueba de Wilcoxon para datos pareados, y la prueba de la χ^2 para datos no pareados se sustituye por la prueba de McNemar para datos pareados.

La descripción general de los métodos estadísticos debe aparecer de forma clara en la sección «Material y métodos», habitualmente en la parte final. Y cuando se resuman los datos en

la sección «Resultados», hay que especificar los principales métodos estadísticos que se emplearon para analizarlos.

2. Cuantificar los resultados. Es clave presentar los resultados con los indicadores apropiados, y cualquier relación de cifras obtenida de un experimento se puede sustituir por tres valores: el número de componentes o tamaño de la muestra; el valor medio como medida de centralización, que informa de alrededor de qué valor se agrupan los datos; y la desviación estándar, como medida de dispersión que permite saber si este agrupamiento de datos es muy disperso o muy concentrado.

Porque nunca se debe exponer una media sin la medida de variabilidad o precisión de la media: la desviación estándar (DE) se usa para conocer la variabilidad entre individuos, y el error estándar de la media (EE) para conocer la variabilidad de la muestra. Así pues, la DE y el EE son dos aspectos distintos que no cuantifican lo mismo: el EE de la media no cuantifica la variabilidad de las observaciones, como hace la DE, sino la precisión con la que una media muestral estima la verdadera media poblacional. Por tanto, es la DE y no el EE de la media lo que debe usarse al resumir los datos. Confundir el EE con la DE es un error (o una trampa), dado que en el EE de la media la variabilidad es siempre menor que en la DE. Si un autor indica el EE y el tamaño de la muestra (n), el lector puede calcular la DE mediante la fórmula: $DE = EE \times \sqrt{n}$.

Debe quedar claro qué medida de variabilidad se presenta, por lo que el signo \pm debe ser sustituido preferentemente por las expresiones DE o EE, o bien dicho aspecto debe especificarse claramente en la sección «Material y métodos». Porque la expresión $9,3 \pm 2,5$ no tiene el mismo valor si se presenta como 9,3 (DE 2,5) que como 9,3 (EE 2,5).

Pero antes de nada, al cuantificar los resultados, conviene tener muy claro lo que no debe hacerse en la presentación de datos numéricos (tabla 3).

3. Conocer el significado del valor de p. Porque el principal objetivo de las pruebas estadísticas es responder a la pregunta: ¿cuál es la probabilidad de que la diferencia observada se deba al azar?

Lo primero es realizar una prueba de hipótesis: se plantea una hipótesis nula (H_0), es decir, $A = B$, y una hipótesis alternativa (H_1), es decir, $A \neq B$. De acuerdo con el resultado del estudio, aceptaremos o, por el contrario, rechazaremos la hipótesis nula. Lo segundo es cuantificar la diferencia observada entre las dos muestras (A y B) mediante un número denominado estadístico de la prueba (p. ej., la t de Student). Cuanto mayor sea la diferencia entre las muestras, mayor será el valor del estadístico de la prueba. Lo tercero es calcular el nivel de

Tabla 3. Errores que deben evitarse en la presentación de datos numéricos

1. Presentar demasiados datos: simplificar es la clave
2. Presentar sólo los datos estadísticamente significativos: lo no significativo puede ser importante
3. Confundir al expresar los datos con cifras o letras: es preferible escribirlos con cifras, pero entre el uno y el diez pueden expresarse también en letras
4. Presentar los datos con una precisión excesiva: la precisión debe ser adecuada, no excesiva
5. Olvidar el denominador en los datos: principalmente en porcentajes, tasas, cocientes, etc.
6. Usar el signo \pm sin indicar a qué medida precede: desviación estándar, error estándar o intervalo de confianza
7. Confundir al hacer un uso incorrecto de los guiones: prestar especial atención a los intervalos, en los que a veces es preferible prescindir del guion, e indicar «de -3 a 12%» en lugar de «-3 -12%»
8. No indicar todas las unidades de todos los datos
9. No presentar todos los datos de todos los sujetos: es importante el flujograma de seguimiento de los sujetos
10. Uso incorrecto de palabras con significado técnico: en estadística, al César lo que es del César y a Dios lo que es de Dios

significación (el valor de p). El valor de p es la probabilidad de equivocarse al afirmar que existe una diferencia verdadera. Una $p < 0,05$ indica que, en menos de 5 veces de cada 100 que repitiéramos el mismo estudio, nuestro resultado sería falso o se debería al azar. La significación estadística a partir de $p > 0,05$ es una convención, pero es también de interés la información que se obtiene de los resultados con una p algo mayor de 0,05. Cuanto mayor sea la p , más fuerte será la evidencia a favor de la hipótesis nula.

Muchos piensan que el valor p es la probabilidad de cometer un error. Pero, obviamente, hay dos formas de cometer un error a partir de los datos. Puede afirmarse que el tratamiento tiene un efecto cuando en realidad no lo tiene y también puede negarse que tenga un efecto que en realidad sí tiene. El valor de p sólo cuantifica la probabilidad de cometer el error tipo I o alfa: afirmar erróneamente que el tratamiento tiene un efecto cuando en realidad no lo tiene (falso positivo). El valor de p no da información alguna sobre la probabilidad de cometer el error tipo II o beta: concluir que el tratamiento no tiene efecto cuando en realidad lo tiene (falso negativo). En la tabla 4 se esquematizan estos tipos de error y los riesgos asociados en un contraste de hipótesis.

¿Cuándo elegir una p con una o dos colas? Generalmente, es conveniente elegir una p con dos colas (A puede ser mejor o peor que B). La p con una cola se utilizará cuando consideremos que A nunca podrá ser peor que B. Esto es importante, ya que, por ejemplo, una $p = 0,04$ con una cola es estadísticamente significativa, mientras que la equivalencia con dos colas ($p = 0,08$) no lo sería.

4. No depender exclusivamente de la p . Porque contar sólo con el uso de los valores de p como prueba para comprobar la hipótesis estadística no nos transmite toda la información cuantitativa importante. Los niveles de significación (valores de p) se suelen dar a conocer para demostrar que la diferencia observada en relación con un patrón (la hipótesis nula) tendría muy pocas probabilidades de ocurrir meramente por azar; sin embargo, dar a conocer tan sólo el valor p de una prueba de significación estadística de las diferencias escatima información.

Las típicas aseveraciones $p > 0,05$ o $p < 0,05$ dan poca información sobre los resultados del estudio y se basan en el consenso arbitrario de utilizar el nivel de significación estadística del 5% para definir dos posibles resultados: no significativo ($p > 0,05$) o significativo ($p < 0,05$). Siempre que sea posible, hay que anotar los valores exactos de p (es más informativo), aunque para simplificar la información se puede expresar la significación estadística en tres intervalos ($p < 0,05$, $p < 0,01$ y $p < 0,001$).

En cualquier caso, hay que hacer una distinción clara entre la significación estadística y la significación clínica: un resultado significativo no indica necesariamente un efecto real y, por el contrario, un resultado no significativo no indica ausencia de efecto, sino sólo que con los datos actuales no existe evidencia del efecto. Debe haber cierta flexibilidad en la interpretación de la significación estadística de los test. Usamos generalmente un punto de corte de 0,05, pero valores de $p = 0,04$ (significativo) y $p = 0,06$ (no significativo) no son muy diferentes del anterior, pese a la interpretación tan diferente. Calificar cualquier valor de $p > 0,05$ co-

Tabla 4. Tipos de error y riesgos asociados en un contraste de hipótesis

		Realidad	
		Existe diferencia (H_0 falsa)	No existe diferencia (H_0 verdadera)
Resultado de la prueba	Diferencia significativa (rechazo de H_0)	No error	Error tipo I (riesgo α)
	Diferencia no significativa (no rechazo de H_0)	Error tipo II (riesgo β)	No error

mo «no significativo» no es recomendable, ya que puede enmascarar resultados con sugestivo efecto si se describe el valor exacto de *p*.

Los intervalos de confianza (IC) constituyen una forma más informativa de presentar las pruebas de significación que un simple valor de *p*. Es más útil presentar los valores muestrales como estimaciones de los resultados que se obtendrían si se estudiara toda la población. La falta de precisión de un valor muestral (p. ej., la media), que depende tanto de la variabilidad del factor que se investiga como del tamaño limitado del estudio, puede ilustrarse muy bien mediante un IC. Un IC nos lleva de un solo valor estimado (p. ej., la media muestral, la diferencia entre medias muestrales, etc.) a un recorrido de valores que se consideran plausibles para la población. El tamaño muestral influye en la magnitud del EE y éste, a su vez, influye en la amplitud del IC. Frente a la interpretación de la significación estadística que sólo indica la probabilidad de que la diferencia encontrada se deba al azar, el IC traduce la incertidumbre en valores reales, por lo que facilita entender su significado clínico. Por ello, en la medicina basada en pruebas se utiliza a menudo el aforismo de «muera la *p*, viva el IC».

Puede ser preferible utilizar la media con un IC razonable de incertidumbre, habitualmente del 95% (que comprende dos EE por debajo y por encima de la media): por ejemplo, 8,2 (IC del 95%: 6,3-10,1). Esta forma de presentar los datos resulta particularmente útil si los resultados del estudio son «negativos», porque, a diferencia del término «no significativo» de los test de significación, el IC refleja el amplio intervalo de efectos que son compatibles con los datos observados.

¿Qué es la precisión estadística? Es la inversa de la variancia (a mayor precisión, menor variancia), que aumenta al aumentar el tamaño de la muestra. Una manera de medir la precisión es mediante el IC: cuanto más amplio es el IC, menor es la precisión, es decir, hay menor confianza en que el resultado no se deba al azar y puede inferirse a la población.

5. Proporcionar los detalles del proceso de aleatorización. La asignación aleatoria de las intervenciones en los ensayos clínicos pretende conseguir que todos los sujetos participantes tengan la misma probabilidad de recibir cada uno de los tratamientos. Existen diferentes métodos de aleatorización; los más usados son los siguientes: 1) asignación aleatoria simple; 2) aleatorización en bloques balanceados; 3) aleatorización estratificada, y 4) aleatorización en conglomerados (grupos o *clusters*). En general, puede decirse que la aleatorización por bloques es preferible a la simple.

6. Proporcionar los detalles del proceso de enmascaramiento (cegamiento). Debe describirse diferenciando si se trata de simple ciego (tratamiento desconocido por el paciente), doble ciego (también desconocido para el médico), triple ciego (también para el monitor) o cuádruple ciego (también en el análisis estadístico).

7. Especificar el número de observaciones e indicar las pérdidas. Es importante considerar las pérdidas prealeatorización (por negarse los pacientes a entrar en el estudio, o bien por ser considerados no elegibles de acuerdo con los criterios de inclusión/exclusión) y postaleatorización (abandonos y retiradas de pacientes durante el estudio). Conviene detallar bien el diagrama de flujo del estudio: valorar a los pacientes que no realizan el tratamiento inicialmente asignado y a los que se pierden en el seguimiento. Una pérdida en el seguimiento superior al 20% compromete la validez de un estudio.

8. Especificar el programa estadístico y las referencias sobre estadística. Hay que concretar las referencias sobre el programa estadístico informático utilizado (licencia disponible, indicando la versión), así como las principales referencias sobre el diseño del estudio y los métodos estadísticos empleados, especialmente en los estudios estadísticos sofisticados, complejos o de uso poco frecuente. Existen múltiples pruebas estadísticas, y es importante conocer y especificar las que se utilizan (tabla 5).

Tabla 5. Pruebas estadísticas según el objetivo y la variable de respuesta*

Objetivo		Variable de respuesta			
		Cuantitativa normal	Ordinal o cuantitativa no normal	Binomial	Tiempo de supervivencia
Comparar dos grupos	Independientes	t de Student	U de Mann-Whitney	Prueba de la χ^2 , prueba de Fisher	<i>Log-rank test</i> , Mantel-Haenszel
	Apareados	t de Student para muestras pareadas	Prueba de Wilcoxon	Prueba de McNemar	Modelos de fragilidad
Comparar tres o más grupos	Independientes	Análisis de la variancia (ANOVA)	Prueba de Kruskal-Wallis	Prueba de la χ^2	Riesgos proporcionales de Cox
	Apareados	Análisis de la variancia (ANOVA)	Prueba de Friedman	Q de Cochran	Modelos de fragilidad
Evaluar la fuerza de la asociación		Correlación de Pearson	Correlación de Spearman		Riesgos proporcionales de Cox
Predecir el valor de una variable en función de otra		Regresión lineal	Regresión no paramétrica	Regresión logística simple	Riesgos proporcionales de Cox
Describir la relación entre una variable de respuesta y diversas variables predictoras		Regresión lineal múltiple		Regresión logística múltiple	Riesgos proporcionales de Cox

*Adaptada de Jiménez¹.

Se ha observado que este hecho, junto con la colaboración en el artículo de expertos en estadística, se relaciona con una mejora en la metodología de los trabajos. Según su grado de participación en el diseño del trabajo y/o el análisis y la interpretación de los datos, los expertos en estadística podrían constar como autores, o bien ser mencionados en la sección «Agradecimientos».

9. Definir los términos y evitar el uso no técnico de términos estadísticos. Porque la ciencia es precisa, y la estadística aún más. Por ello, la notificación estadística es una parte importante de un artículo y no deben utilizarse términos estadísticos con su significado común, tales como «al azar», «normal», «significativo», «correlación» o «muestra».

«Al azar» implica el empleo de un método de aleatorización. «Normal» es un conjunto de distribución de probabilidad descrita por una fórmula específica, también referido como distribución gaussiana. «Significativo» quiere decir que el resultado de la prueba formal de una hipótesis estadística se encuentra fuera de una zona determinada de antemano. «Correlación» es la manera específica de medir la asociación entre dos o más variables. «Muestra» suele referirse a una observación o conjunto de observaciones reunidas de una manera bien definida.

Por otro lado, los símbolos más usuales en la expresión de términos estadísticos están basados en las normas dictadas por la Organización Internacional de Estandarización. Pero como la realización tipográfica no es sencilla en cuanto a los símbolos estadísticos, se permite la utilización de abreviaturas alternativas; así, la desviación estándar se puede definir como DE (en lugar de s) y el error estándar de la media como EE (en lugar de Sx).

10. Controlar los errores de la investigación clínica. Algunas reglas básicas pueden ayudar al lector a detectar posibles errores, dado que la mayoría de los errores se deben al mal uso de técnicas elementales que los investigadores deberían aprender a utilizar por sí mismos.

Un resultado estadísticamente significativo (la consabida p) no supone ninguna garantía de que sea un resultado válido, que pueda inferirse a la población general, si no hemos cuidado el control de los errores estadísticos que se producen en los estudios. Centremos este aspecto en el «patrón de oro» de la investigación: el ensayo clínico. La muestra de pacientes incluidos en un ensayo clínico proviene de una población de muestreo, y ésta, a su vez, proviene de la población objetivo. Se pretende que los resultados obtenidos en la muestra de pacientes puedan ser inferidos (generalizados) a la población objetivo. Por ello deben controlarse los errores aleatorios y los errores sistemáticos (o sesgos). El objetivo es controlar al máximo dichos errores, para conseguir: 1) que la muestra seleccionada sea representativa de la población: para ello deben controlarse los errores aleatorios, mediante el muestreo al azar y el aumento del tamaño muestral; y 2) la asignación de los pacientes de la muestra a grupos de tratamiento debe dar lugar a una distribución similar de factores pronóstico, para lo cual deben controlarse los errores sistemáticos (sesgos). El grado de ausencia de sesgos determinará la validez del estudio, y debe saberse que existen dos tipos de sesgos principales: el sesgo de selección de pacientes en el muestreo y en la asignación de los grupos de tratamiento, y el sesgo de evaluación de los resultados (sobre ellos profundizaremos más adelante).

Medidas comunes en epidemiología: medidas de riesgo y de impacto

El esfuerzo para familiarizarnos con conceptos epidemiológicos y estadísticos básicos nos ayudará como clínicos y como investigadores. Los resultados de los estudios epidemiológicos deben ser expresados en forma de medidas de salud o enfermedad, con las que indicamos frecuencias, diferencias, asociación, riesgo o impacto. La forma de presentar los resultados dependerá del tipo de estudio realizado, pero sobre todo de las características de la variable o las variables de interés. A efectos de presentación y análisis, podemos distinguir entre variables continuas (variables cuantitativas medidas en escalas de razones o intervalos), variables ordinales (variables cuantitativas o cualitativas medidas en escalas ordinales) y variables discretas (variables cualitativas medidas en escalas nominales, dicotómicas o politómicas). En función de los tipos de variables implicados, podremos recurrir a distintas medidas epidemiológicas.

En epidemiología, el escenario más simple lo constituye el estudio de dos variables discretas dicotómicas. Este escenario corresponde al supuesto habitual de estudio de la asociación entre presencia-ausencia de un determinado factor de exposición y presencia-ausencia de enfermedad. Para analizar esta relación contamos con una serie de medidas de frecuencia, riesgo e impacto. Otro escenario habitual es el que evalúa la asociación entre una variable discreta y otra continua. Este escenario corresponde a estudios que evalúan la repercusión de un factor de exposición (p. ej., tratamiento frente a placebo) sobre un efecto cuantificable en un rango continuo de valores (p. ej., presión arterial); en estos estudios la presentación de resultados se basará en las diferencias de medidas de tendencia (media, mediana) y dispersión de la variable continua entre grupos.

Incidencia y prevalencia

Incidencia y prevalencia son las expresiones de frecuencia de enfermedad más utilizadas en la literatura médica, y es importante diferenciarlas:

- La incidencia es el número de nuevos casos que ha habido durante un determinado intervalo de tiempo, dividido por el tamaño de la población en riesgo al comienzo del intervalo.

Esta información habitualmente se obtiene de estudios de cohortes y se expresa en forma de tasas.

- La prevalencia es el número de individuos con una enfermedad o característica en un determinado punto en el tiempo, dividido por la población en riesgo en ese momento. Se calcula generalmente a partir de estudios transversales y se expresa como una proporción.

Ambas medidas expresan información complementaria. Una enfermedad con alta incidencia y con alta mortalidad tendrá una baja prevalencia en la población. Alternativamente, una enfermedad con una baja incidencia tendrá una alta prevalencia si su tasa de mortalidad es baja. El efecto de la mortalidad en la prevalencia puede repercutir en las características de las muestras seleccionadas para participar en un estudio, ya que la población susceptible de entrar en un estudio con casos prevalentes estará formada por una selección de pacientes con un mejor pronóstico y una menor presencia de factores de riesgo que la población identificada en un estudio con casos incidentes.

Medidas de riesgo

Mientras que la incidencia y la prevalencia expresan frecuencia de enfermedad en grupos concretos de pacientes, otras medidas permiten comparar las diferencias de frecuencia entre dos grupos distintos o estimar el incremento de riesgo asociado a una exposición concreta (tabla 6). Las principales medidas de riesgo son el riesgo relativo y la *odds ratio* (razón de ventajas):

- El riesgo relativo (RR) se calcula dividiendo la incidencia en el grupo de sujetos expuestos a un determinado factor de riesgo o protección entre la incidencia en el grupo no expuesto. Sólo puede ser calculado en estudios de seguimiento, y mide la fuerza de la asociación entre exposición y enfermedad. Adopta valores entre 0 e infinito, menores de 1 para factores de protección y mayores para factores de riesgo; un RR de 1 supone el valor nulo (el riesgo en los dos grupos es el mismo); cuanto más se aleje el valor de 1, por arriba o por abajo, mayor será la fuerza de la asociación.

Cuando el seguimiento realizado a los sujetos incluidos en un estudio es heterogéneo, en vez de considerar la incidencia acumulada, para el cálculo de riesgos, se recurre a la densidad de incidencia, en la que cada sujeto es considerado en función del tiempo que es seguido. Para calcular la densidad de incidencia, se realiza un cociente donde en el numerador se sitúa el número de casos nuevos y en el denominador el número de sujetos seguidos por el número de unidades de tiempo (habitualmente años) que es seguido cada uno de ellos. En este caso, el riesgo relativo se estima a partir de la razón de densidades de incidencia (RDI) entre grupos.

- En los estudios sin seguimiento longitudinal (estudios de casos y controles), como no es posible calcular la incidencia, no puede calcularse el riesgo relativo. Por ello, para estimar el riesgo se calcula la *odds ratio* (OR), que compara la *odds* de exposición (probabilidad de estar expuesto a un factor de riesgo dividida entre su complementario) en el grupo con enfermedad y la *odds* de exposición en el grupo control sin enfermedad; la OR es el resultado de dividir ambas *odds*. La interpretación de la OR es similar a la del RR: un valor de 1 supone el valor nulo, los valores menores de 1 indican una disminución del riesgo, y los mayores, un aumento del riesgo. Hay que tener en cuenta que, sólo cuando la enfermedad estudiada es muy poco frecuente, el RR y la OR ofrecen valores similares.

Medidas de impacto

Aunque con las medidas anteriores podemos estimar el riesgo que genera un factor de exposición sobre un efecto o una enfermedad, esas medidas no nos informan del impacto que dicha exposición puede originar en el conjunto de casos existentes en una población. Esta información podemos extraerla de otras medidas, como la diferencia de riesgos o la proporción atribuible (tabla 6).

Tabla 6. Análisis de los estudios de cohortes. Medidas de riesgo e impacto

Estudios con seguimiento acumulativo			
Factor de exposición			
	Cohorte expuesta	Cohorte no expuesta	
Enfermos	a	b	N_1
No enfermos	c	d	N_0
	M_1	M_0	T
Incidencia acumulada (IA) en expuestos	$IA_1 = a / M_1$		
Incidencia acumulada en no expuestos	$IA_0 = b / M_0$		
Riesgo relativo (RR)	$RR = \frac{IA_1 \text{ expuestos}}{IA_0 \text{ no expuestos}} = \frac{a/M_1}{b/M_0}$		
Estudios con datos personas-tiempo			
Factor de exposición			
	Cohorte expuesta	Cohorte no expuesta	
Casos nuevos (enfermos)	a_i	a_0	
Total personas-tiempo	L_1	L_0	T
Personas-tiempo = personas seguidas por tiempo de seguimiento de cada una			
Densidad de incidencia (DI)			
Razón de densidades de incidencia (RDI)	$RDI = \frac{DI_1 \text{ expuestos}}{DI_0 \text{ no expuestos}} = \frac{a_i/L_1}{a_0/L_0}$		
Medidas de impacto			
Diferencia de riesgos (DR)	$DR = IA_1 - IA_0 = \frac{a}{M_1} - \frac{b}{M_0}$		
Proporción atribuible (PA)	$PA = \frac{IA_1 - IA_0}{IA_1} = 1 - \frac{IA_0}{IA_1} = 1 - \frac{1}{RR} = \frac{RR - 1}{RR}$		
PA poblacional (PAP)	$PAP = PA \times FE_{\text{casos}} = \frac{RR - 1}{RR} \times \frac{a}{N_1}$		
FE_{casos} : fracción de exposición en los casos (proporción de casos expuestos)			

Ambas medidas son estimadores del efecto absoluto que ejerce la exposición sobre la incidencia de un suceso en el grupo expuesto o en la población total. Se usan para evaluar la relevancia clínica o sanitaria de una exposición y nos informan sobre el porcentaje de incidencia que se reduciría si se eliminara el referido factor de exposición. Son, por consiguiente, muy útiles, tanto en la clínica como en salud pública, para cuantificar el posible impacto de diferentes medidas de intervención.

- La diferencia de riesgos (DR) se calcula restando de la incidencia en el grupo expuesto al factor de riesgo la incidencia en el grupo no expuesto. Adopta valores entre 0 y 1 (entre 0 y 100 si se expresa en porcentajes), constituyendo el 0 el valor nulo de ausencia de diferencias. La DR ofrece información independiente del riesgo relativo y puede variar, entre

distintos grupos de pacientes, en función del riesgo propio de cada grupo. Así, podemos encontrarnos que factores que muestran un riesgo relativo muy alto apenas presenten modificaciones en la diferencia de riesgos porque el riesgo en la población (al margen de la contribución de dicho factor) sea muy bajo.

- La proporción atribuible (PA), también conocida como riesgo atribuible, fracción etiológica, fracción atribuible o población de riesgo atribuible, se define como la proporción de casos nuevos de la enfermedad, en el grupo de sujetos expuestos, que son atribuibles al factor de riesgo de interés. Se calcula dividiendo la diferencia de riesgos, anteriormente calculada, entre la incidencia en el grupo expuesto. Una extensión de esta medida es la proporción atribuible poblacional, que extiende la proporción de casos nuevos a toda la población, esto es, al conjunto de sujetos tanto expuestos como no expuestos.
- Los resultados de los ensayos clínicos suelen reflejar el efecto beneficioso de intervenciones terapéuticas que reducen el riesgo en el grupo expuesto. Por ello, la diferencia de riesgos en este supuesto, conocida como reducción absoluta del riesgo (RAR), se calcula en sentido contrario, restando del riesgo en el grupo control el riesgo en el grupo de intervención.
- Otra medida de impacto aplicable a estos estudios, de gran interés clínico, es el número necesario a tratar (NNT), que corresponde al inverso de la RAR ($1/RAR$) y nos informa del número de pacientes que deberían ser tratados con la intervención terapéutica para que un caso se viera beneficiado, evitándose un evento desfavorable.

Tipos de estudios epidemiológicos: entre la causalidad y la «evidencia»

Diseño de los estudios epidemiológicos

La epidemiología estudia la distribución de las enfermedades o los acontecimientos relacionados con la salud, y trata de identificar sus causas, su evolución natural y el papel que desempeñan determinados factores de riesgo o protección. Para ello, emplea una metodología basada en el método científico; el punto de partida es la identificación de una pregunta de investigación inspirada en el conocimiento actual sobre un determinado problema de salud. La contestación a esta pregunta requiere la formulación de una hipótesis de trabajo, que ha de contrastarse en un estudio llevado a cabo en una muestra de población. Generalmente, la pregunta de investigación de un estudio epidemiológico intenta identificar la relación entre un factor de estudio o exposición, que puede ser de riesgo o protección, y una variable de efecto o respuesta, habitualmente curación o enfermedad. La estrategia que sigamos para medir ambos parámetros, especialmente la secuencia temporal en que son valorados, condicionará el grado de evidencia que finalmente obtendremos sobre dicha relación.

Podemos clasificar los estudios epidemiológicos en función de distintos criterios⁵: el objetivo de la pregunta de investigación (descriptivos o analíticos), las características del diseño (direccionalidad, tipo de muestreo y temporalidad), la existencia de intervención y las unidades de estudio.

- Según el objetivo general de nuestra pregunta de investigación, distinguiremos dos grupos de estudios: los descriptivos y los analíticos. Los descriptivos tratan de contestar preguntas del tipo ¿qué pasa?, ¿a quién?, ¿cuántos?, ¿dónde? o ¿cuándo?, mientras que los analíticos intentan responder a cuestiones como ¿por qué? ¿cómo evitarlo? o ¿sirvió para algo? La principal diferencia entre ambos tipos de estudio radica en que los analíticos requieren, para contestar sus preguntas, la existencia de grupos de comparación, con los que contrastar la hipótesis formulada.

- El diseño de los estudios epidemiológicos puede ser catalogado en relación con tres características: direccionalidad, tipo de muestreo y temporalidad.
 - La direccionalidad se refiere al orden en el que se investigan la variable de exposición (factor de riesgo o protección) y la variable efecto o respuesta (enfermedad, curación). Un estudio tendrá una direccionalidad anterógrada cuando la medición de la exposición anteceda a la del efecto, retrógrada cuando se mida primero el efecto y posteriormente se busque la exposición, y simultánea cuando ambas variables se determinen en el mismo momento. Sólo la direccionalidad anterógrada es capaz de garantizar que la ocurrencia de la exposición preceda en el tiempo a la del efecto, lo que constituye uno de los criterios clásicos de causalidad.
 - También podemos diferenciar los estudios en función del tipo de muestreo; así, los sujetos de estudio podrán ser seleccionados por la presencia o ausencia de un factor de exposición (de riesgo o protección), de una variable de resultado o efecto (enfermedad), o bien a partir de otros criterios de selección ajenos a ambos factores.
 - La tercera característica de diferenciación en cuanto al diseño de un estudio es la temporalidad. La temporalidad se refiere a la relación entre el tiempo en el que el investigador realiza el estudio y la ocurrencia temporal del factor de exposición y del efecto. La temporalidad es histórica cuando la exposición y el efecto han ocurrido antes de iniciarse el estudio, concurrente cuando la exposición y el efecto tienen lugar durante el proceso de investigación y mixta cuando el factor de exposición es histórico y el efecto concurrente. A la hora de garantizar la validez de la información recogida en un estudio, es la temporalidad concurrente la que ofrece un mayor grado de confianza.

La utilización de los términos «prospectivo» y «retrospectivo» para referirse al tipo de diseño de un estudio es un motivo frecuente de confusión. Algunos autores emplean estos términos para expresar direccionalidad, mientras que otros indican con ellos temporalidad. Asimismo, ambos términos se han empleado para denominar tipos concretos de estudio, como los estudios de cohortes (prospectivos) y de casos y controles (retrospectivos). También se han empleado para referirse a la estrategia de análisis estadístico, en función de que las pruebas de contraste de hipótesis se hayan definido antes de conocer los datos o durante el análisis. Considerando estas cuestiones, parece prudente evitar el empleo de estos términos a la hora de describir el diseño de un estudio.
- Como hemos visto, el principal interés de los estudios epidemiológicos es explorar la relación entre un factor de exposición y un efecto. Con frecuencia, al establecer esta relación interfieren otros factores que no podemos controlar. Por ello, la mejor manera de controlar la repercusión de una exposición es introducirla de forma controlada, de modo que desencadenemos una secuencia de acontecimientos que den lugar o no al efecto buscado. Si como investigadores somos meros espectadores y no modificamos el factor de exposición, estaremos realizando estudios observacionales; si por el contrario lo manipulamos, estaremos llevando a cabo estudios experimentales o de intervención. Aunque los estudios experimentales son los que mayor grado de evidencia aportan, determinadas preguntas de investigación no pueden ser contestadas con este diseño, ya que la intervención no sería ética o viable.
- Finalmente, los estudios se diferencian en función de las unidades de estudio consideradas. Así, podemos distinguir entre estudios realizados sobre sujetos individuales y estudios realizados sobre grupos de población. Estos últimos caracterizan una serie de diseños diferenciados; para los estudios descriptivos, los denominados «estudios ecológicos», y para los de intervención, los conocidos como «ensayos comunitarios».

En el algoritmo de la figura 1 se recogen, en función de sus características, los diversos tipos de estudios epidemiológicos⁵, representándose cada uno de ellos tanto más a la derecha y abajo cuanto mayor es el grado de evidencia que aportan. En el algoritmo no aparecen recogidos los estudios ecológicos y los ensayos comunitarios.

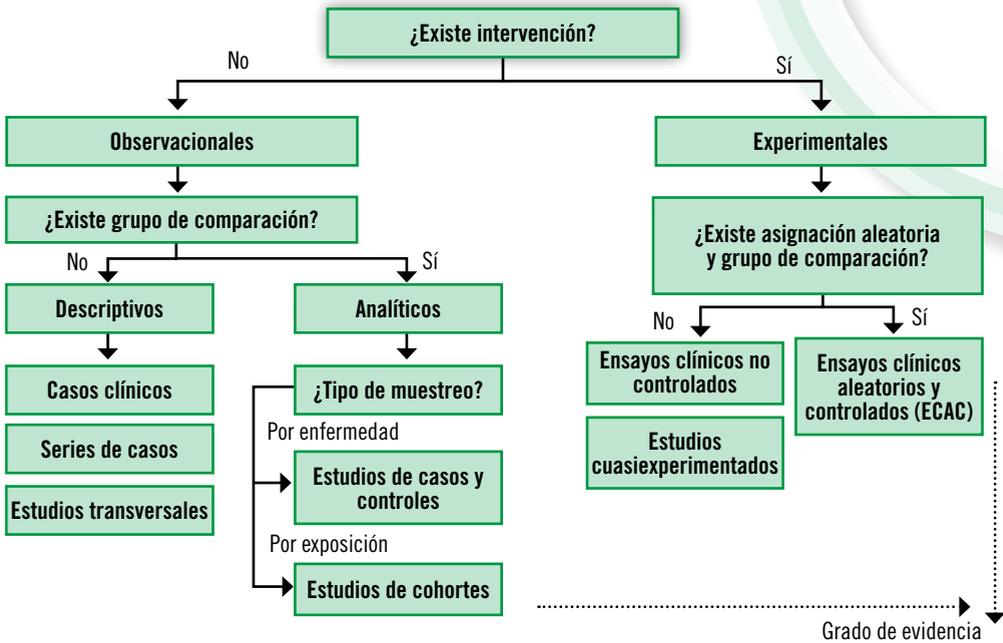


Figura 1. Tipos de estudios epidemiológicos y nivel de «evidencia»

- Estudios transversales⁶. Son estudios observacionales, descriptivos y simultáneos (no direccionales). Estudian la población total o una muestra significativa de ella, en un punto del tiempo o en un periodo corto, y buscan la presencia o ausencia de factores de riesgo (exposición) y enfermedad (efecto). También son denominados estudios «de prevalencia», ya que permiten cuantificar esta medida de enfermedad.

Los estudios transversales son muy útiles en la planificación sanitaria, ya que permiten describir la distribución de las enfermedades y los factores de riesgo, así como sugerir hipótesis que puedan comprobarse en otros estudios analíticos. Tienen como ventajas que son fáciles de ejecutar, relativamente poco costosos y fácilmente repetibles. Entre sus inconvenientes podemos destacar que son estudios sujetos a múltiples sesgos, sobre todo de selección (es primordial que la muestra elegida sea representativa), no sirven para la investigación causal por carecer de direccionalidad, no permiten estimar incidencias ni riesgos, y son poco útiles para estudiar enfermedades raras o de corta duración.

- Estudios de cohortes⁷. Son estudios observacionales, analíticos, generalmente anterógrados, de temporalidad concurrente o mixta, en los que se muestrea en relación con la exposición. Se conocen también como estudios «de seguimiento», «de proyección», «de incidencia» o «prospectivos».

El punto de partida de estos estudios es la selección de dos grupos de sujetos, diferenciados según tengan o no un determinado factor de exposición, que se siguen a lo largo del tiempo para determinar la proporción de cada grupo que desarrolla la enfermedad objeto de estudio. Todos los sujetos tienen que estar libres de dicha enfermedad al inicio del estudio, pero ser susceptibles de poder padecerla.

Las principales ventajas de los estudios de cohortes es que permiten calcular incidencias y estimar riesgos de forma más fiable (tabla 6), son eficientes para estudiar factores de exposición poco frecuentes, establecen la secuencia real de eventos (de exposición a enfermedad) y controlan mejor la existencia de sesgos. Como inconvenientes, cabe citar que son difíciles de ejecutar, de coste elevado y difíciles de reproducir, que no son factibles para

Tabla 7. Análisis de estudios de casos y controles. Odds ratio

		Efecto (enfermedad)		
		Casos	Controles	
Exposición presente		a	b	N_1
Exposición ausente		c	d	N_0
		M_1	M_0	T

Odds de enfermedad en expuestos = a/b
Odds de enfermedad en no expuestos = c/d

Odds ratio (OR) $OR = \frac{a/c}{b/d} = \frac{a \times d}{b \times c}$

enfermedades raras o con un periodo de inducción largo (requerirían un gran tamaño muestral y un largo seguimiento), y que se ven influidos por cambios en las técnicas de medición a lo largo del tiempo o por desigualdades en la colaboración y seguimiento de las cohortes.

- Estudios de casos y controles⁸. Son estudios observacionales, analíticos, en los que el muestreo se hace en relación con la enfermedad o el efecto, de direccionalidad retrógrada y temporalidad mixta. Se conocen también como estudios «de casos y testigos» o «retrospectivos».

En estos estudios se parte de dos grupos de sujetos: unos que tienen la enfermedad (casos) y otros que carecen de ella (controles), y se investiga en ambos grupos si en el pasado hubo determinadas características (factores de exposición o riesgo) presumiblemente relacionadas con la enfermedad.

Si los pasos previos se han realizado correctamente, a partir de los resultados podremos realizar estimaciones de riesgo, calculando la OR (tabla 7).

Las ventajas de los estudios de casos y controles es que permiten estudiar enfermedades poco frecuentes, son de fácil ejecución, son muy eficientes (menor coste por unidad de estudio), tienen una buena capacidad para generar hipótesis y, si están bien diseñados y ejecutados, pueden alcanzar una validez cercana a la de los estudios de cohortes. Sin embargo, los estudios de casos y controles también conllevan algunos inconvenientes: no permiten calcular medidas de frecuencia (incidencia o prevalencia), ofrecen una peor estimación del riesgo que los estudios de cohortes, al no garantizar la secuencia temporal de la relación entre el factor de exposición y la enfermedad plantean dificultades para el manejo de variables tiempo-dependientes (hábitos dietéticos y terapéuticos) y presentan una mayor susceptibilidad a los sesgos.

- Ensayos clínicos aleatorizados y controlados⁹. Son estudios de intervención, analíticos, anterógrados, de temporalidad concurrente y muestreo de cohorte cerrada con control de la exposición. Requieren la asignación aleatoria de la intervención que introduce el factor de estudio y la existencia de grupo de comparación. Los ensayos clínicos permiten evaluar los beneficios de los tratamientos o programas de intervención, ponderando sus riesgos y efectos adversos. Con ellos podemos controlar los sesgos existentes en otros tipos de estudio y obtener información sólida sobre la relación causal entre una exposición (intervención) y un efecto (inferencia causal).

Para analizar sus resultados, debemos contrastar la comparabilidad de los grupos antes de la intervención y cuantificar las principales medidas de efecto e impacto (tabla 8). Los ensayos clínicos tienen importantes ventajas: son los estudios que aportan mayor control sobre los distintos factores implicados en el diseño, por lo que son repetibles y compara-

Tabla 8. Análisis de los ensayos clínicos

Tasa de respuesta en el grupo de intervención	$P_i = \frac{\text{n}^\circ \text{ eventos grupo intervención}}{\text{total sujetos grupo intervención}}$
Tasa de respuesta en el grupo control	$P_c = \frac{\text{n}^\circ \text{ eventos grupo control}}{\text{total sujetos grupo control}}$
Reducción relativa del riesgo	$RRR = \frac{P_c - P_i}{P_c}$
Reducción absoluta del riesgo	$RAR = P_c - P_i$
Número necesario a tratar	$NNT = \frac{1}{RAR}$

bles; a menudo ofrecen el único diseño posible para contestar a determinadas preguntas de investigación; además, proporcionan la mejor evidencia sobre inferencia causal y generalización de los resultados. Como contrapartida, también tienen diversos inconvenientes: son estudios costosos en tiempo y dinero, pueden verse limitados por problemas éticos o de factibilidad, y en ocasiones la aplicación de sus resultados se ve cuestionada porque la muestra de estudio no representa a la población en la que se desea aplicar esos resultados.

Causalidad

Como hemos comentado previamente, el principal objetivo de la mayoría de los estudios epidemiológicos es identificar si existe alguna relación entre una determinada exposición (factor o factores de riesgo o protección) y un efecto (enfermedad, curación, etc.). Habitualmente, la hipótesis de interés es si la exposición está causalmente relacionada con el efecto. Esta causalidad puede definirse operativamente de la siguiente manera: «Una exposición es causa de un efecto si un determinado nivel de exposición da lugar a un efecto (o nivel de efecto) distinto del que hubiera existido sin dicha exposición (o nivel de exposición)». En esta definición, el término «exposición» está precedido de las palabras «nivel de» para incluir el criterio comparativo en el concepto de causalidad. Antes de atribuir una relación causal a una exposición, debe estar bien definido con respecto a qué, o lo que es lo mismo, cuál es el nivel de referencia.

La existencia de una asociación, por fuerte que sea, entre una exposición y un efecto no implica causalidad; cualquier relación causal que se sugiera debe ser coherente con el estado actual del conocimiento o tener algún fundamento fisiopatológico que la respalde (plausibilidad biológica).

Teniendo en cuenta que la evaluación causal de un determinado factor de exposición se basa en los resultados de estudios epidemiológicos en grupos de población, la evidencia causal dependerá de la validez interna y externa de dichos estudios (ausencia de sesgos). Otros criterios que considerar en la causalidad son la fuerza de la asociación (efecto estadística y clínicamente significativo), el gradiente biológico (a mayor exposición, mayor efecto), la concordancia externa (con otros estudios) y la direccionalidad. Este último crite-

rio resulta fundamental, ya que la exposición debe anteceder al efecto. Sólo los diseños epidemiológicos de direccionalidad anterógrada (ensayos clínicos y estudios de cohortes) permiten controlar dicha relación temporal; por ello están situados en posición preferente dentro de las escalas de jerarquización de la evidencia.

Porque causalidad y «evidencia» están interrelacionadas, y un factor que debe tenerse en cuenta es el manejo de los errores en los estudios epidemiológicos.

Errores en los estudios epidemiológicos: evitar el error aleatorio y el error sistemático

Todo estudio epidemiológico ha de ser entendido como un ejercicio de medición, en el que debe perseguirse la exactitud, esto es, la estimación del parámetro deseado con el mínimo error. Existen dos posibles fuentes de error: el error aleatorio y el error sistemático. Un estudio es preciso cuando carece de error aleatorio y se considera válido cuando carece de error sistemático.

Evitar el error aleatorio

Por error aleatorio (debido al azar) entendemos el error ligado al propio proceso de medición, es decir, a la falta de precisión. Esta falta de precisión o error aleatorio tendrá un valor distinto, e impredecible dentro de un rango, en cada una de las repeticiones de la medición. La variabilidad de una medición será, por tanto, un indicador de su precisión; cuanto más amplia sea dicha variabilidad, menor será su precisión. La variabilidad puede deberse a distintos factores: variabilidad biológica, imprecisión del instrumento de medida, inexperiencia del evaluador, etc.

Los errores aleatorios no suelen afectar a la validez interna de los estudios (no alteran la dirección de los resultados), pero sí limitan su potencia. En un estudio epidemiológico, la principal manera de reducir el error aleatorio consiste en aumentar el tamaño de la muestra. De este modo, minimizamos el riesgo de obtener resultados distorsionados por azar, ya que la repetición de la medición tenderá a producir resultados distintos pero cercanos al valor verdadero del parámetro que se quiere medir.

Asumiendo que la repetición de un experimento con un mismo número de pacientes puede dar resultados diferentes por mero azar, la inferencia estadística nos permite cuantificar el rango de error, a partir de medidas de dispersión de los resultados obtenidos y del tamaño muestral (error estándar). A menor dispersión de los resultados y mayor tamaño muestral, menor rango de error habrá.

- Sea cual sea el parámetro elegido para expresar los resultados de un estudio, el rango de error de nuestra estimación puede expresarse como un intervalo de confianza, situado entre un valor por debajo y otro por encima del resultado obtenido. Cuanto mayor grado de confianza queramos atribuir a nuestra estimación de error, más amplio será el rango de dicho intervalo. Habitualmente empleamos el intervalo de confianza del 95%, cuya interpretación es la siguiente: «Si repetimos 100 veces un experimento en las mismas condiciones y con igual número de sujetos, el verdadero valor del parámetro poblacional que queremos estimar se encontrará incluido dentro de 95 intervalos de confianza de los 100 construidos a

partir de los resultados de cada experimento». Similar interpretación tendrían intervalos de confianza más exigentes.

La presentación de la incertidumbre, sobre el tamaño del efecto obtenido en un estudio, en forma de intervalos de confianza es el método preferido en las principales revistas científicas. A diferencia del nivel de significación obtenido en un contraste de hipótesis (la p), que sólo expresa la probabilidad de que la diferencia encontrada se deba al azar (error tipo I o de falso positivo), el intervalo de confianza refleja la incertidumbre del resultado en las propias unidades de medida del efecto que debe evaluarse, lo que facilita la interpretación de su relevancia clínica. Por otra parte, el intervalo de confianza permite juzgar la significación estadística del resultado, ya que sólo será significativo cuando no esté comprendido en su rango el valor nulo de ausencia de efecto (p. ej., 0 para diferencias de medias o riesgos, 1 para medidas de riesgo).

- También podemos medir el error aleatorio mediante el abordaje clásico que ofrece el contraste de hipótesis. El objetivo del contraste de hipótesis es permitir generalizaciones de los resultados de nuestra muestra a la población de la que procede. Cuando el resultado de un estudio en el que se comparan varias alternativas muestra diferencias, el siguiente paso es estimar si dichas diferencias corresponden a diferencias reales en la población o pueden deberse al azar. Si la probabilidad de que el resultado obtenido se deba al azar es muy baja, podremos asumir que corresponde a una diferencia real en la población, aunque siempre existirá una cierta probabilidad de error. Esta probabilidad será tanto más baja cuanto mayor sea la diferencia encontrada y mayor sea el tamaño muestral.

El contraste de hipótesis se plantea habitualmente bajo el supuesto de dos hipótesis contrapuestas: una hipótesis nula conservadora que sostiene la ausencia de diferencias (p. ej., el porcentaje de curación con el nuevo tratamiento es igual al que se obtiene con un tratamiento clásico) frente a una hipótesis alternativa novedosa que defiende la existencia de diferencias (p. ej., el porcentaje de curación con el nuevo tratamiento es mayor). Para poder aceptar la hipótesis alternativa debe rechazarse la hipótesis nula, lo que se consigue cuando la diferencia o efecto encontrado en el estudio no parece deberse al azar. Diversos procedimientos estadísticos (basados en los datos del estudio y en el tamaño muestral) nos permiten calcular la probabilidad de que dicho resultado se produzca por azar. Por convención, si esta probabilidad es menor del 5% ($p < 0,05$) se considera que no se debe al azar; en ese caso se rechaza la hipótesis nula y se acepta la alternativa. Esta decisión se tomará asumiendo siempre cierto margen de error, equivalente a la probabilidad anteriormente calculada (error tipo I; α) y expresada en el valor p de significación. No obstante, el valor de p no ilustra la magnitud de los resultados obtenidos, como sí hace el intervalo de confianza; su nivel de significación, arbitrariamente elegido, se verá muy influido por el tamaño de la muestra, al margen de cuál sea el efecto encontrado; por ello, ofrecerá poca información sobre la relevancia clínica de los resultados.

En ocasiones, las diferencias encontradas en un estudio no alcanzan el nivel de significación del 5%, porque la probabilidad de que los resultados sean debidos al azar es mayor ($p \geq 0,05$). En estos casos, el contraste de hipótesis no permite rechazar la hipótesis nula. Sin embargo, hemos podido encontrarnos cifras mayores del 5% porque la diferencia real en la población sea pequeña, en cuyo caso no cometeremos ningún error, o bien porque hayamos obtenido una diferencia muy pequeña por azar, en cuyo caso sí estaremos cometiendo un error (error tipo II, β). Este error será tanto más probable cuanto menor sea el tamaño muestral de nuestro estudio, lo que producirá un aumento del error aleatorio que se traducirá en el cálculo de probabilidades. Como consecuencia de ello, nuestro estudio tendrá poca potencia para encontrar diferencias, lo que se expresará como el complementario de la probabilidad de error tipo II ($1 - \beta$). Distintos procedimientos estadísticos permiten calcular a partir de la hipótesis alternativa (en la que estarán reflejadas las diferencias clínicamente relevantes) cuál es la potencia de un estudio; si ésta no alcanza al menos un 80% (error tipo II menor del 20%), la potencia de nuestro estudio será insuficiente, y deberemos plantearnos una ampliación de la muestra.

Evitar el error sistemático

Por error sistemático entendemos la falta de validez en la medición. También se lo denomina sesgo (*bias*). La validez se refiere al concepto de «medir lo que se quiere realmente medir». La validez es, por tanto, la carencia de error sistemático. El error sistemático o sesgo es atribuible a errores metodológicos que, a diferencia del error aleatorio, no se reducen aumentando el tamaño muestral.

En la mayoría de los estudios de investigación clínica o epidemiológica es difícil obtener la suficiente información como para cuantificar apropiadamente la magnitud de los sesgos. Sin embargo, sí es posible conocer la dirección de estos errores, lo que reviste extraordinaria importancia para el investigador, pues si encontramos una fuerte asociación entre variables es crucial conocer si potenciales sesgos pudieron modificar al alza o a la baja dicha estimación.

En la validez de un estudio de investigación se suele distinguir entre dos conceptos: la validez interna y la validez externa. Se entiende que hay validez interna cuando los resultados del estudio son válidos entre los propios sujetos del estudio. Por validez externa se entiende la capacidad de generalización de los resultados, es decir, la capacidad de extender y aplicar las inferencias obtenidas a partir de ellos a una población diana (o población objetivo, también llamada externa) de referencia. Aunque una muestra mayor consigue una mayor precisión en las estimaciones disminuyendo el error aleatorio, la representatividad de una muestra no puede definirse en un sentido estadístico, sino que requiere un juicio racional sobre el parecido de la muestra a la población. Por otra parte, la existencia de validez interna es un requisito previo para que pueda darse la validez externa.

Podemos clasificar los sesgos en sesgos de selección, sesgos de información (o clasificación) y sesgos de análisis (confusión e interacción).

- Existe un sesgo de selección cuando se produce una distorsión del efecto estimado como consecuencia de errores en el proceso de selección de los sujetos. Si la elección de los grupos que se van a comparar no es apropiada, se producirá un sesgo de selección. Es un sesgo que se comete con frecuencia en los estudios de casos y controles, y ante el que conviene estar especialmente alerta.
- Los sesgos de información o clasificación consisten en la distorsión del efecto estimado por errores en la medición y/o clasificación de los sujetos en una o más variables. Algunos ejemplos son: instrumentos de medición no válidos o no calibrados, criterios diagnósticos inapropiados o cambiantes, pérdidas de información, seguimiento desigual entre los grupos de estudio, incoherencia entre los entrevistadores, ausencia de interpretación ciega de la exposición o del efecto, etc. La influencia de estos errores en la clasificación del nivel de exposición o en el efecto puede distorsionar nuestros resultados, especialmente cuando los errores son diferenciales. Entendemos por error de clasificación diferencial el producido cuando el criterio de clasificación en uno de los grupos que se comparan es diferente del criterio utilizado en el otro grupo. El efecto de este error suele traducirse en que los sujetos de un grupo son clasificados incorrectamente y los del otro correctamente. La estimación del efecto puede verse aumentada o disminuida, produciéndose una distorsión en contra o a favor de la hipótesis nula. El error de clasificación no diferencial es el cometido cuando se clasifica incorrectamente a los dos grupos por igual. La clasificación de exposición o de enfermedad es errónea para iguales proporciones de sujetos en los dos grupos de comparación. Este error tiende a producir una subestimación del efecto, esto es, una distorsión a favor de la hipótesis nula, y en consecuencia genera resultados más conservadores. Por ello, el error no diferencial ocasiona menos problemas que el error diferencial.
- Mientras que los sesgos de selección o clasificación sólo pueden preverse y corregirse en la fase de diseño, otros errores pueden ser controlados en la fase de análisis. De estos sesgos de análisis, los principales son la confusión y la interacción.

– Cuando el efecto estimado de la relación exposición-enfermedad está distorsionado por la presencia de algún factor extraño en el análisis de los datos, decimos que existe confusión. El estimador del efecto de interés está distorsionado por estar mezclado con el efecto producido por un factor extraño, que denominamos factor (o variable) de confusión. La presencia de tal factor de confusión altera los resultados y puede producir un aumento o una disminución del efecto, o incluso cambiar la dirección del mismo.

Para que un factor sea considerado factor de confusión debe cumplir ciertos requisitos: 1) ser factor de riesgo independiente de la enfermedad tanto para los expuestos como para los no expuestos en la población; 2) estar asociado al factor de exposición que estudiamos en la población de donde provienen los casos, y 3) no ser una consecuencia de dicha exposición, es decir, no ser un paso intermedio en la cadena o secuencia de causalidad entre la exposición y la enfermedad. Un sesgo de confusión puede corregirse antes de iniciar el estudio, durante el proceso de diseño, empleando técnicas de restricción, pero también puede corregirse durante el análisis, mediante técnicas de ajuste, fundamentalmente el análisis estratificado y el análisis multivariante.

– El otro sesgo de análisis que debemos controlar es el de modificación del efecto o interacción. Este sesgo ocurre cuando el efecto producido por una determinada variable se ve modificado en presencia de una tercera variable, produciéndose efectos de distinta intensidad, en la misma o distinta dirección. Este fenómeno debe ser distinguido del de confusión y también debe ser controlado en el análisis.

Las listas guía de comprobación: un mar de siglas para navegar mejor en metodología y estadística

En nuestra práctica clínica intentamos resolver de la mejor forma posible la ecuación entre «lo deseable, lo posible y lo apropiado» en medicina, una ciencia sembrada de incertidumbre, variabilidad en la práctica clínica y sobrecarga de información. El paradigma de la medicina basada en la evidencia (o en pruebas científicas) intenta ofrecer un poco más de «ciencia» al «arte» de la medicina, un arte que juega continuamente con la probabilidad, y que intenta ayudar a solventar esta solución de continuidad entre información y conocimiento, y aún más, entre conocimiento y práctica. Publicar es algo inherente al trabajo del médico, pero en el actual contexto de exceso de información médica («infoxicación») es necesario disponer de buenas publicaciones, científicamente válidas y éticamente correctas, y para ello necesitamos guías de buena práctica para la elaboración de trabajos (tanto en aspectos formales como éticos) y que sean guías que se divulguen y estén disponibles para editores, autores, revisores y lectores. A estas guías se las denomina «listas guía de comprobación».

Las listas guía de comprobación de artículos son herramientas diseñadas para ayudar a alcanzar ciertos estándares de calidad, ética y rigor científico, al proporcionar un conjunto de recomendaciones que orientan hacia una conducta adecuada. Una edición deficiente de las publicaciones científicas (sean ensayos clínicos u otros tipos de estudios, como estudios de cohortes, de casos y controles, descriptivos o de pruebas diagnósticas) indica un fracaso colectivo del proceso científico y de publicación biomédica: un fracaso de los autores, los revisores, los directores de revistas, las editoriales, los financiadores y el propio sistema de investigación en general.

El propósito de las listas guía es obtener de los autores una exposición transparente de la investigación para que tanto sus lectores como sus evaluadores comprendan su diseño, realiza-

Tabla 9. Declaración CONSORT 2010 para el ensayo clínico aleatorizado			
Sección/tema	Ítem n.º	Ítem de la lista de comprobación	Informado en página n.º
Título y resumen	1a	Identificado como un ensayo aleatorizado en el título	
	1b	Resumen estructurado del diseño, métodos, resultados y conclusiones del ensayo	
Introducción			
Antecedentes y objetivos	2a	Antecedentes científicos y justificación	
	2b	Objetivos específicos o hipótesis	
Métodos			
Diseño del ensayo	3a	Descripción del diseño del ensayo (p. ej., paralelo, factorial), incluida la razón de asignación	
	3b	Cambios importantes en los métodos después de iniciar el ensayo (p. ej., criterios de selección) y su justificación	
Participantes	4a	Criterios de selección de los participantes	
	4b	Procedencia (centros e instituciones) en que se registraron los datos	
Intervenciones	5	Intervenciones para cada grupo con detalles suficientes para permitir la replicación, incluidos cómo y cuándo se administraron realmente	
Resultados	6a	Especificación a priori de las variables de respuesta (o desenlace) principal/es y secundarias, incluyendo cómo y cuándo se evaluaron	
	6b	Cualquier cambio en las variables de respuesta tras el inicio del ensayo, junto con los motivos de la/s modificación/es	
Tamaño muestral	7a	Cómo se determinó el tamaño muestral	
	7b	Si corresponde, explicar cualquier análisis intermedio y las reglas de interrupción	
Aleatorización			
Generación de la secuencia	8a	Método utilizado para generar la secuencia de asignación aleatoria	
	8b	Tipo de aleatorización; detalles de cualquier restricción (como bloques y tamaño de los bloques)	
Mecanismo de ocultación de la asignación	9	Mecanismo utilizado para implementar la secuencia de asignación aleatoria (como contenedores numerados de modo secuencial), describiendo los pasos realizados para ocultar la secuencia hasta que se asignaron las intervenciones	
Implementación	10	Quién generó la secuencia de asignación aleatoria, quién seleccionó a los participantes y quién asignó los participantes a las intervenciones	
Enmascaramiento	11a	Si se realizó, a quién se mantuvo cegado después de asignar las intervenciones (p. ej., participantes, cuidadores, evaluadores del resultado) y de qué modo	
	11b	Si es relevante, descripción de la similitud de las intervenciones	
Métodos estadísticos	12a	Métodos estadísticos utilizados para comparar los grupos en cuanto a la variable de respuesta principal y las secundarias	

Tabla 9. Declaración CONSORT 2010 para el ensayo clínico aleatorizado

Sección/tema	Ítem n.º	Ítem de la lista de comprobación	Informado en página n.º
	12b	Métodos de análisis adicionales, como análisis de subgrupos y análisis ajustados	
Resultados			
Flujo de participantes (se recomienda encarecidamente un diagrama de flujo)	13a	Para cada grupo, el número de participantes que se asignaron aleatoriamente, que recibieron el tratamiento propuesto y que se incluyeron en el análisis principal	
	13b	Para cada grupo, pérdidas y exclusiones después de la aleatorización, junto con los motivos	
Reclutamiento	14a	Fechas que definen los periodos de reclutamiento y de seguimiento	
	14b	Causa de la finalización o de la interrupción del ensayo	
Datos basales	15	Una tabla que muestre las características basales demográficas y clínicas para cada grupo	
Números analizados	16	Para cada grupo, número de participantes (denominador) incluidos en cada análisis y si el análisis se basó en los grupos inicialmente asignados	
Resultados y estimación	17a	Para cada respuesta o resultado final principal y secundario, los resultados para cada grupo, el tamaño del efecto estimado y su precisión (como intervalo de confianza del 95%)	
	17b	Para las respuestas dicotómicas, se recomienda la presentación de los tamaños del efecto tanto absoluto como relativo	
Análisis secundarios	18	Resultados de cualquier otro análisis realizado, incluido el análisis de subgrupos y los análisis ajustados, diferenciando entre los especificados a priori y los exploratorios	
Daños (perjuicios)	19	Todos los daños (perjuicios) o efectos no intencionados en cada grupo (para una orientación específica, véase <i>CONSORT for harms</i>)	
Discusión			
Limitaciones	20	Limitaciones del estudio, abordando las fuentes de posibles sesgos, las de imprecisión y, si procede, la multiplicidad de análisis	
Generalización	21	Posibilidad de generalización (validez externa, aplicabilidad) de los hallazgos del ensayo	
Interpretación	22	Interpretación coherente con los resultados, con balance de beneficios y daños, y considerando otras evidencias relevantes	
Otra información			
Registro	23	Número de registro y nombre del registro de ensayos	
Protocolo	24	Dónde puede accederse al protocolo completo del ensayo, si está disponible	
Financiación	25	Fuentes de financiación y otras ayudas (como suministro de medicamentos), papel de los financiadores	

ción, análisis e interpretación. Estas listas guía conforman un mar de siglas (AGREE, CONSORT, COREQ, QUORUM, MOOSE, PRISMA, REMARK, SQUIRE, STARD, STROBE, TREND, etc.)¹⁰ que definen iniciativas para mejorar la realización y publicación de estudios experimentales, observacionales, de precisión diagnóstica, pronóstico, evaluación económica, guías de práctica clínica, etc. Las más conocidas son la AGREE (Appraisal of Guidelines for REsearch & Evaluation), la CONSORT (Consolidated Standards of Reporting Trials) para ensayos clínicos y la recientemente aparecida PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) para revisiones sistemáticas y metaanálisis, y que reemplaza la QUOROM.

Pues bien, una parte esencial de estas listas guía son las orientaciones que proporcionan para desenvolverse mejor con la metodología y la estadística. En la tabla 9 se ofrece el ejemplo de la declaración CONSORT para ensayos clínicos expuesto.

La mayor utilización de técnicas estadísticas cada vez más complejas en los artículos científicos plantea una serie de temas que conviene considerar para poder sobrevivir a ella, importantes para todos los protagonistas de la publicación científica: editores, autores y lectores.

- A los editores, esa complejidad estadística les dificulta su trabajo de selección editorial. En este sentido, las revistas biomédicas de más prestigio cuentan con un comité de revisores especialistas en metodología y estadística que asesora al comité editorial.
- A los autores se les exige un mayor conocimiento de los diseños y las técnicas de análisis. Este hecho, dada la insuficiencia de contenidos metodológicos en el currículo de medicina, impone la necesidad de constituir equipos multidisciplinares formados por clínicos y ex-

Tabla 10. Errores estadísticos frecuentes en los artículos de investigación biomédica*

1. Presentar datos con precisión innecesaria
2. No presentar todos los datos ni todos los pacientes
3. Dividir datos continuos en categorías ordinales sin explicar el por qué ni el cómo
4. Presentar medias de grupos para datos apareados sin exponer los cambios intraparejas
5. Utilizar incorrectamente los datos descriptivos
6. Utilizar el error estándar de la media (EEM) como estadístico descriptivo o como medida de precisión de una estimación
7. Presentar sólo los valores de p de los resultados
8. No utilizar el intervalo de confianza en cada uno de los estadísticos
9. No confirmar que los datos cumplen las asunciones de las pruebas estadísticas empleadas para analizarlos
10. Utilizar el análisis de regresión lineal sin establecer si la relación es realmente lineal
11. No informar de si se han realizado ajustes para las comparaciones múltiples o de cómo se han hecho
12. Presentar innecesariamente comparaciones estadísticas de la línea basal en los ensayos clínicos
13. No definir «normal» o «anormal» cuando se informa de los resultados de pruebas diagnósticas
14. No explicar cómo se han tratado los valores inciertos (equivocos) de las pruebas diagnósticas al calcular sus características (como sensibilidad y especificidad)
15. Utilizar tablas y figuras sólo para almacenar datos, y no para ayudar a los lectores
16. Utilizar tablas y figuras en las que el mensaje visual no concuerde con los datos del texto
17. Interpretar estudios con resultados no significativos y baja potencia estadística como «negativos», cuando de hecho son no concluyentes
18. No distinguir entre estudios pragmáticos (efectividad) y explicativos (eficacia) al diseñar e interpretar la investigación
19. No presentar los datos en unidades clínicamente útiles
20. Confundir significación estadística con importancia clínica

*Modificada de Jiménez¹.

Tabla 11. Las trampas de la estadística*

1. Realizar comparaciones directas sin el adecuado contraste de hipótesis
2. No ajustar según las diferencias basales entre los dos grupos
3. No especificar claramente el método de aleatorización
4. Operaciones aritméticas inadecuadas con variables cualitativas
5. Uso de una prueba incorrecta para el contexto del trabajo
6. Buscar la prueba estadística que dé el resultado deseado
7. No respetar el protocolo original del estudio. Análisis posterior de subgrupos
8. Realizar contraste de hipótesis unilateral para alcanzar significación estadística
9. No ajustar el efecto producido por valores extremos o eliminarlos inadecuadamente
10. Uso indebido de reflexión y correlación
11. Mal uso del valor de p . Realización de comparaciones múltiples
12. Uso de medidas de efecto relativas en lugar de absolutas

*Modificada de Jiménez¹.

ertos en metodología como mejor estrategia para obtener el máximo rendimiento de la investigación biomédica.

- A los lectores, la mayor utilización de técnicas estadísticas les dificulta el acceso a un porcentaje creciente de originales. Al lector, que no tiene por qué ser un experto en metodología en ciencias de la salud, se le exigen cada vez más conocimientos para poder entender y analizar críticamente la literatura científica que consulta.

La estadística no debe convertirse en una religión, en un instrumento que conduzca a la salvación mediante el ritual de la búsqueda de valores significativos de la p , alcanzando entonces los autores el paraíso prometido de la publicación del trabajo en una revista de alto impacto; la estadística es una herramienta útil para el diseño y la realización de estudios médicos, siempre que se utilice adecuadamente.

Un correcto manejo de la estadística por parte del autor mejorará la calidad científica del artículo y, secundariamente, facilitará que el artículo sea aceptado por el editor y difundido por los lectores. Los textos sobre normas de publicación científica contienen apartados sobre cómo hacer un buen uso de la estadística y cómo evitar su abuso o su mal uso, e incluyen también listas de errores estadísticos; en este sentido, puede resultar muy útil empezar tomando conciencia de los peligros: los principales errores estadísticos en los que se incurre (tabla 10) y las trampas de la estadística¹¹ (tabla 11).

Agradecimientos

Al Grupo de Trabajo de Pediatría Basada en la Evidencia y, en especial, a los Dres. Carlos Ochoa y Manuel Molina, por el continuo aprendizaje durante estos más de diez años de trabajo conjunto.

BIBLIOGRAFÍA

1. Jiménez J. Presentación de resultados estadísticos y elaboración de tablas. En: Jiménez Villa J, Argimon Pallàs JM, Martín Zurro A, Vilardell Tarrés M. Publicación científica biomédica: cómo escribir y publicar un artículo de investigación. Barcelona: Elsevier España, 2010; 57-77. Libro de consulta imprescindible en español para la buena redacción de un artículo de investigación.
2. González de Dios J. Conocimientos prácticos para elaborar un artículo científico (IV): manejo de la estadística. Rev Pediatr Esp. 1998; 54: 340-345. La estadística aplicada en la redacción de un artículo científico, con un enfoque práctico y sencillo.
3. González de Dios J, González Muñoz M, Alonso A, Aleixandre R. Comunicación científica (X). Conocimientos básicos para elaborar un artículo científico (5): los aspectos estadísticos (más que números). Acta Pediatr Esp. 2014; 72(3): 63-70. Similar a la anterior, esta publicación forma parte de una serie de artículos sobre comunicación científica.
4. Emerson JD, Colditz GA. Use of statistical analysis in The New England Journal of Medicine. N Engl J Med. 1983; 309(12): 709-713. Artículo clave para conocer mejor las características de las categorías estadísticas y el concepto de umbral estadístico.
5. Molina Arias M, Ochoa Sangrador C. Tipos de estudios epidemiológicos. Evid Pediatr. 2013; 9(3): 53.
6. Molina Arias M, Ochoa Sangrador C. Estudios observacionales (I). Estudios transversales. Medidas de frecuencia. Técnicas de muestreo. Evid Pediatr. 2013; 9(4): 72.
7. Molina Arias M, Ochoa Sangrador C. Estudios observacionales (II). Estudios de cohortes. Evid Pediatr. 2014; 10(1): 14.
8. Molina Arias M, Ochoa Sangrador C. Estudios observacionales (III). Estudios de casos y controles. Evid Pediatr. 2014; 10(2): 33.
9. Molina Arias M, Ochoa Sangrador C. Ensayo clínico (IV). Resultados. Variables y medidas de efecto e impacto. Evid Pediatr. 2015; 11(2): 33. Artículos sencillos y prácticos, dentro de la serie Fundamentos en Medicina Basada en la Evidencia.
10. González de Dios J, González Muñoz M, Alonso A, Aleixandre R. Comunicación científica (XVIII). Conocimientos básicos para leer (y escribir) un artículo científico (5): lista de comprobación de documentos. Acta Pediatr Esp. 2014; 72(11): e389-e392. Artículo esencial para entender las listas de comprobación de documentos; forma parte de una serie de artículos sobre comunicación científica.
11. Molina Arias M. Las trampas de la estadística. Rev Pediatr Aten Primaria. 2014; 16(62): 181-186. La serie Lectura Crítica en Pequeñas Dosis nos ofrece una visión asequible y divertida de la estadística y de la lectura crítica de documentos científicos.