



Metodología

La significancia estadística y los intervalos de confianza: ¿qué me indican y cómo puedo interpretarlos?

Statistical significance and confidence intervals: What do they tell me and how can I interpret them?

Martha A. Sánchez-Rodríguez¹

¹ Profesora de Tiempo Completo Titular C de la Carrera de Química Farmacéutico Biológica, FES Zaragoza, UNAM.

INTRODUCCIÓN

Cuando nos introducimos a la lectura de los artículos científicos nos enfrentamos a diferentes situaciones, desde la comprensión del tema hasta el análisis de los resultados. La lectura de un artículo científico debe realizarse en el contexto de los resultados, es decir, lo que nos debe interesar es la información nueva que aporta, por lo que hay que enfocarse en ellos, habiendo revisado previamente la sección de material y métodos para comprender qué fue lo que se hizo y cómo se hizo, además del análisis estadístico llevado a cabo.

La finalidad de la sección de resultados del artículo científico es mostrar lo encontrado en una investigación proporcionando la mayor información de la forma más sencilla, clara, concisa, precisa, con secuencia lógica y estética, es decir, una buena presentación visual para una lectura rápida y fácil. Los resultados de un estudio pueden presentarse de forma narrativa, en cuadros o figuras que preferentemente deben ser autoexplicativos y autosuficientes,^{1,2} teniendo las figuras diferentes modalidades como son gráficas, esquemas, dibujos, o fotografías, por mencionar las más comunes. Los cuadros o figuras autoexplicativos favorecen el análisis preliminar de los datos sin hacer referencia al texto completo, pero se corre el riesgo de una mala interpretación,¹ por lo que debemos hacer una lectura detallada de este apartado, entendiendo la información incluyendo los datos estadísticos proporcionados.

De los resultados cuantitativos o cualitativos se debe indicar el número de observaciones y el recorrido de los datos observados, tener las medidas adecuadas de error o incertidumbre, y especificar las pruebas estadísticas utilizadas para analizar esos resultados;³ de ahí que para entender esta sección es necesario conocer algunos conceptos estadísticos esenciales.

En este trabajo revisaremos brevemente los conceptos de significancia estadística, también llamado valor de p , y los intervalos de confianza, su interpretación y aplicación en ejemplos concretos.

Correspondencia: Martha A. Sánchez-Rodríguez

E.mail: masanrod@yahoo.com.mx

Artículo recibido: 14 de abril de 2021

Artículo aceptado: 26 de abril 2021

Sánchez-Rodríguez MA. La significancia estadística y los intervalos de confianza: ¿qué me indican y cómo puedo interpretarlos? *CyRS*. 2021; 3(1):74-82

DOI: <https://doi.org/10.22201/fesz.26831422e.2021.3.1.8>

SIGNIFICANCIA ESTADÍSTICA Y VALOR DE p

La prueba de significancia es una técnica de inferencia estadística para juzgar si una característica, que se supone cumple una población, es compatible con lo observado en una muestra de ella,⁴ y comprende el conjunto de pruebas estadísticas que permiten tomar una decisión referente a una población, examinando una muestra de esa población, por lo que es el fundamento de toda la estadística del análisis de los datos y nos ayuda a comprender la interpretación.

En este artículo no es la intención tratar el tema detalladamente, pues lo pueden consultar en cualquier libro de estadística, sino retomar los puntos básicos para entender el significado del valor de p . Para ello requerimos recordar que para establecer si una prueba es estadísticamente significativa es necesario el planteamiento de dos hipótesis: una llamada *Hipótesis nula* (H_0), que establece que no existe evidencia para creer que haya una diferencia entre una muestra y su población, y que, si se encuentra es debida al azar. Se plantea generalmente para rechazarla y es la base formal para examinar la significancia estadística; y otra denominada *Hipótesis alterna* (H_a o H_1), que establece la diferencia entre la muestra y la población, y que generalmente es la hipótesis de trabajo. Una regla importante es que ambas hipótesis deben ser mutuamente excluyentes.^{5,6}

Posterior al planteamiento de las hipótesis, se debe calcular un parámetro estadístico para rechazar o no H_0 , esto es, seleccionar la prueba estadística adecuada a lo que se quiere probar, una diferencia o no entre grupos o una asociación entre variables. El parámetro estadístico, también llamado estadígrafo, servirá para establecer si H_0 se rechaza o no se rechaza con un nivel de significancia llamado α . Este nivel de significancia es un valor fijo establecido para cada estadígrafo al hacer el cálculo y se traduce en la probabilidad de no rechazar H_0 cuando es verdadera o error tipo I, y representa el azar.

Por convención estadística, el límite que separa lo posible de lo no posible es 5 en 100, es decir, que se acepta el 5% de posibilidad de equivocación para rechazar H_0 siendo verdadera y que esta equivocación sea debida al azar, esto representado en valores de probabilidad se expresa como p

$=0.05$.^{7,8} De ahí que el procedimiento del cálculo de la prueba estadística proporciona un valor de probabilidad denominado valor de p (p value), que es la probabilidad de que se presente un valor del estadístico más alejado de H_0 que el observado, por lo que se conoce también como nivel de significancia.^{4,9} Ambos términos, α y p , son llamados nivel de significancia, sin embargo, son diferentes ya que α es un valor fijo establecido antes del análisis y p es un valor de probabilidad obtenido del análisis de los datos,¹⁰ y lo que interpretamos en los resultados de un análisis son los valores de p .

En la interpretación, si se rechaza H_0 , se puede decir que las muestras observadas difieren significativamente de los resultados esperados, por lo que la prueba calculada es estadísticamente significativa. Si H_0 no se rechaza se dice que los datos sobre los cuales se basa la prueba no proporcionan evidencia suficiente que provoque el rechazo, por lo que la prueba no es estadísticamente significativa.^{6,9} El valor de p es, por lo tanto, la probabilidad de que el resultado obtenido se deba al azar si la hipótesis nula es cierta, siendo responsable de la variación de los resultados. Como el valor de $p = 0.05$ es el límite de equivocación en la decisión estadística, un valor de $p < 0.05$ es el término común para señalar que una prueba es **estadísticamente significativa**.^{6,8}

Un valor de p pequeño, < 0.05 , representa una probabilidad muy baja de que el valor de la prueba estadística se hubiera generado con los datos si H_0 fuera verdadera, por ello es estadísticamente significativa. Un valor de $p = 0.001$ se considera muy significativo, y nos dice que lo observado en un estudio es muy raro (sólo en 0.1% de los casos), por lo que mientras más pequeño es el valor de p ,^{6,10} más significativa es la prueba estadística calculada. Los valores de p significativos pueden indicarse de manera puntual (es lo más recomendable), poniendo el valor exacto obtenido en el análisis estadístico, o globalizado redondeando al valor más cercano de acuerdo con los valores de corte establecidos: < 0.05 , < 0.01 , < 0.001 o < 0.0001 .

Contrariamente, un valor de $p > 0.05$, no significativo, no demuestra que H_0 es verdadera, sino que puede ser debido a un pequeño tamaño de muestra o una gran variabilidad de los datos, por lo que



Table 1 Baseline characteristics of women by study group

Parameter	Premenopausal women (n = 101)	Postmenopausal women (n = 101)	p value
Biochemical -hematologic parameters			
Hemoglobin (g/dL)	14.0 ± 1.5	14.6 ± 1.4	0.002
Hematocrit (%)	43 ± 3.9	45 ± 3.9	0.003
Erythrocytes (X10 ¹² /L)	4.6 ± 0.5	4.8 ± 0.7	0.034
Total leukocytes (X10 ⁹ /L)	6431 ± 1573	6071 ± 1192	0.070
Glucose (mmol/L)	5.6 ± 2.5	5.2 ± 1.8	0.773
Urea (mmol/L)	9.6 ± 2.1	10.4 ± 2.5	0.085
Uric acid (µmol/L)	274 ± 77	280 ± 77	0.735
Creatinine (µmol/L)	72 ± 15	72 ± 12	0.848
Cholesterol (mmol/L)	5.6 ± 0.9	5.9 ± 1.1	0.057
Triglycerides (mmol/L)	2.0 ± 1.0	2.1 ± 1.1	0.552
HDL-c (mmol/L)	1.5 ± 0.4	1.6 ± 0.4	0.832
Estrogen (pg/mL)	101.7 ± 71.5	12.3 ± 6.2	<0.0001
FSH (mIU/mL)	10.7 ± 12.2	55.2 ± 26.1	<0.0001

Quantitative data show means ± standard deviation. HDL-c high density lipoprotein cholesterol, BMI body mass index. The p value was determined with Student's t test

Figura 1. Cuadro de resultados de una investigación con los valores de p puntuales y diferentes interpretaciones. Marco azul, diferencia estadísticamente significativa ($p < 0.05$); marco rojo, sin diferencia estadísticamente significativa, y marco amarillo, diferencia altamente significativa. Tomado de: Sánchez-Rodríguez (2017).¹¹

Table 2: Oxidative stress markers in healthy and osteoporosis subjects

	Healthy (n = 50)	Osteoporosis (n = 44)
Lipoperóxidos (LPO) (µmol/L)	0.258 ± 0.09	0.296 ± 0.14
Superoxide dismutase (SOD) (UI/L)	167 ± 10.7	165 ± 7.4
Glutathione peroxidase (GPx) (UI/L)	7039 ± 2724	5399 ± 2359*
Total antioxidant status (TAS) (mmol/L)	1.03 ± 0.21	0.96 ± 0.18
SOD/GPx ratio	0.028 ± 0.01	0.036 ± 0.01†
Mean values ± SD; t-test, *p < 0.01, †p < 0.05.		

Figura 2. Cuadro de resultados con valores de p señalados con identificadores (círculos rojos) indicando diferentes niveles de significancia y la prueba estadística utilizada (marco rosa). Tomado de Sánchez-Rodríguez (2007).¹²

se interpreta como que no hay suficiente evidencia para rechazar H_0 .^{6,8} Por lo general, en un estudio no se reportan los valores de $p > 0.05$, por lo que, si no aparece el valor de p , se sobreentiende que el análisis llevado a cabo no mostró un parámetro estadísticamente significativo.

Los valores de p deben estar en la presentación de los resultados que fueron comparados o con los que se pretende demostrar una asociación. Se pueden encontrar de forma narrativa, en cuadros o figuras (gráficas). En la figura 1 se muestra la presentación de resultados de un artículo en un cuadro, nos indica que el análisis se realizó con la prueba t de Student (señalada con la flecha roja) y se incluyó el valor de p puntual (círculo verde). Cuando se reportan los resultados de esta manera, nos corresponde identificar cuáles son estadísticamente significativos y cuáles no, recordando que cualquier prueba estadística es significativa cuando el valor de $p < 0.05$. En este contexto, en la figura se observa que los parámetros de la fórmula roja son estadísticamente diferentes entre los grupos (marco azul), la cuenta de leucocitos y la química sanguínea no muestran ninguna diferencia (marco rojo), y los niveles hormonales son muy diferentes, ya que el valor de $p < 0.0001$ (marco amarillo).

Cuando en el cuadro de resultados no se incluyen los valores de p puntuales, se deben poner marcas señalando los resultados que muestran una diferencia significativa y, en muchas ocasiones, los identificadores son diferentes dependiendo del nivel de significancia (círculos rojos), además de indicar la prueba estadística utilizada (marco rosa mexicano) como se aprecia en la figura 2.

En el caso de las gráficas, se estila que se coloquen marcas en las representaciones de barras o líneas mostrando la diferencia entre subgrupos o momentos. Igual que en los cuadros, los valores de p pueden ser puntuales o globales.

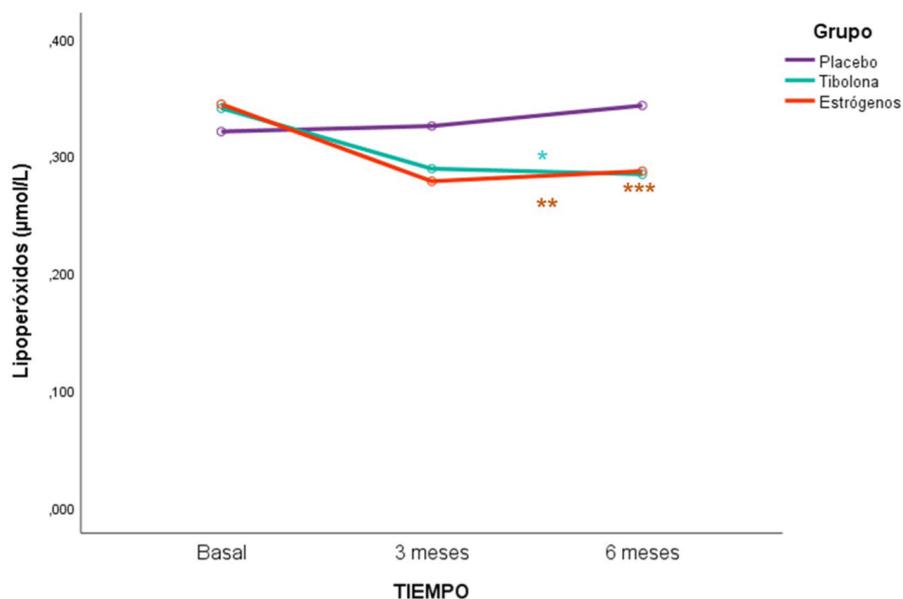


Figura 14. Media de los niveles de lipoperóxidos en los grupos de estudio, al inicio 3 y 6 meses de tratamiento. ANOVA de medidas repetidas con prueba de Bonferroni como post hoc. Tibolona, basal – 6 meses $p = 0.038$; estrógenos, basal – 3 meses $p = 0.006$ y basal – 6 meses $p = 0.008$.

Figura 3. Gráfica de líneas del efecto de tres tratamientos sobre un biomarcador de estrés oxidativo a través del tiempo indicando los valores de p puntuales (círculos azules) y las pruebas estadísticas utilizadas (marco rosa mexicano). Tomado de: Ruiz-Rodríguez (2018).¹³

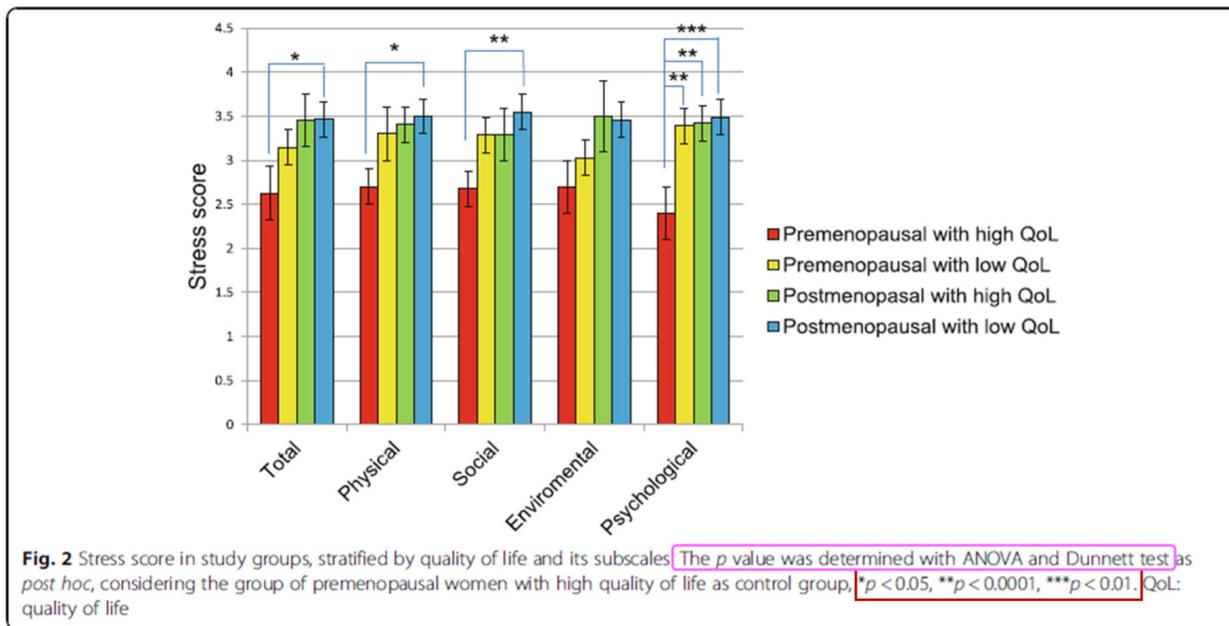


Fig. 2 Stress score in study groups, stratified by quality of life and its subscales. The p value was determined with ANOVA and Dunnett test as post hoc, considering the group of premenopausal women with high quality of life as control group. $p < 0.05$, $p < 0.0001$, $p < 0.01$. QoL: quality of life

Figura 4. Gráfica de barras con barras de error que muestra las comparaciones estadísticamente significativas entre subgrupos, las pruebas estadísticas usadas (marco rosa) y los valores de p globales (marco rojo). Tomado de: Sánchez-Rodríguez (2017).¹¹



En la figura 3 observamos una gráfica de líneas del efecto de tres tratamientos sobre un biomarcador de estrés oxidativo, las marcas en la gráfica representan cambios estadísticamente significativos a través del tiempo, los cuáles son reportados como valores puntuales (círculos azules) junto con las pruebas estadísticas empleadas (marco rosa mexicano).

En la figura 4 se muestra una gráfica de barras con barras de error en donde con líneas se indica la comparación que resultó estadísticamente significativa, entre subgrupos, señalada con asteriscos. En la descripción de la figura se incluye el análisis estadístico utilizado (marco rosa mexicano) y el nivel de significancia globalizado (marco rojo) para los diferentes asteriscos.

Actualmente existe una discusión con relación a la información proporcionada por el valor de p , ya que está limitado a una probabilidad de que los resultados observados sean producto del azar, por lo que se hacen algunas sugerencias en la interpretación:^{6,7,10} (i) tomar con cautela un valor cercano a 0.05; (ii) no considerar sólo el valor de p para la toma de decisiones clínicas; (iii) proporcionar, en la medida de lo posible, los valores de p puntuales, ya que son más informativos; (iv) utilizar otros cálculos para la interpretación correcta, como la construcción de intervalos de confianza, como veremos más adelante.

INTERVALO DE CONFIANZA (IC)

Es un intervalo con un límite inferior y uno superior, dentro del cual asumimos que un resultado es verdadero en la población estudiada con un nivel de confianza especificado, generalmente el 95%, por lo que se conoce como intervalo de confianza al 95% ($IC_{95\%}$).

Tiene su fundamento también en la teoría de la inferencia estadística; de ahí que el $IC_{95\%}$ es una herramienta para determinar el rango de valores en los que es probable que se encuentren los parámetros de la población objetivo con una confianza del 95%; es decir, si se toma un número infinito de muestras de una población de un mismo tamaño y se calcula el IC, el parámetro verdadero de la población estará incluido en ese intervalo el 95% de las veces, dejando un 5% al azar de que el parámetro verdadero esté fuera de ese intervalo.⁶

Entonces, el IC es una medida de variabilidad o dispersión que controla el error de muestreo o error muestral, considerando el efecto de obtener resultados diferentes debidos al azar en la selección de muestras distintas, pero no puede controlar los errores no muestrales como los sesgos en el diseño o de medición en el estudio.¹⁴ En este sentido, la amplitud del IC indica la precisión de una estimación y depende de la variabilidad de los datos y el número de observaciones en la muestra.⁶

El IC completa la información de la prueba de significancia y ayuda a interpretarla, ya que indica sobre los valores aceptables del parámetro. Cuando no se rechaza H_0 , el IC permite distinguir entre mucha (IC muy amplio) y poca variación (IC estrecho).⁴

Este cálculo puede obtenerse de una gran cantidad de mediciones estadísticas o estimaciones puntuales como: media, proporciones, diferencias de medias y proporciones, razones de riesgos, razón de momios, pruebas de concordancia, pruebas de confiabilidad diagnóstica, etc., en cuyo caso siempre indicará la variabilidad en los datos, pero en algunas estimaciones es tan necesario que no se puede hacer una correcta interpretación si no se obtuvo el IC, como en las razones de riesgos y de momios.

En general, para el cálculo del IC se requiere de una estimación puntual, una medida de variación (error estándar) y un coeficiente de confianza. La estimación puntual es la medida obtenida de la muestra (media, proporción, razón de momios, diferencia de medias, etc.); el error estándar (EE) es la medida de la variación de los datos y se obtiene de diferente manera dependiendo de la estimación puntual calculada considerando el tamaño muestral; y el coeficiente de confianza es comúnmente el valor de z (parámetro de la distribución normal estándar) al 95% de confianza, llamado z_{α} , que en una estimación bilateral ($z_{\alpha/2}$) siempre será 1.96. Para calcular el $IC_{95\%}$ se debe restar y sumar a la estimación puntual el valor del coeficiente de confianza multiplicado por el EE.^{6,14}

$$IC_{(95\%)} = \text{Estimación puntual} \pm 1.96EE$$

Characteristic	Premenopausal Women (n = 145)	Postmenopausal Women (n = 170)
Anxiety	37 (26%, 19–33%)	55 (32%, 25–39%)
Depressive mood	35 (24%, 17–31%)	51 (30%, 23–37%)
Insomnia	74 (51%, 43–59%)	110 (65%, 58–72%) ^c
Smokers (> 2 cigarettes/d)	28 (19%, 13–25%)	15 (9%, 5–13%) ^c
Caffeinated beverages intake (> 2 cups/d)	49 (34%, 26–42%)	48 (28%, 21–35%)
Alcohol intake (> 2 glasses/d)	7 (5%, 1–9%)	7 (4%, 1–7%)
Sedentary (<30 min/d of physical activity)	90 (62%, 58–66%)	99 (58%, 51–65%)

categorical data show frequency, percentage and 95% confidence interval. ^cchi square test ^cp < 0.05.

Figura 5. Ejemplo de presentación de resultados de intervalo de confianza para proporciones en un cuadro de resultados. Marco rojo, IC_{95%} no debidos al azar; marco verde, IC_{95%} tal vez debidos al azar. Tomado de: Sánchez-Rodríguez (2019)¹⁶

De acuerdo con este cálculo, una gran dispersión de los datos y un pequeño tamaño de muestra mostrará un gran EE y un IC_{95%} muy amplio, indicando que los datos de la muestra son insuficientes para precisar la estimación del efecto en la población total, por lo que debe ser interpretado con cautela, independientemente de si son o no estadísticamente significativos;⁶ debido a que el IC describe la variabilidad entre la medida obtenida en una muestra y la medida real de la población y su precisión depende del tamaño de la muestra para un mismo nivel de confianza.¹⁵

Los IC_{95%} se pueden presentar como datos, ya sea en la narrativa de los resultados o en un cuadro (Figuras 5 y 6), o de forma esquemática en algunos tipos de gráficas como las *forest plot* (Figura 7). Podemos apreciar que no importa que parámetro se esté reportando, el valor de la estimación puntual está dentro del intervalo.

Para la interpretación del IC de las proporciones tomaremos algunos resultados de la figura 5. En este cuadro encontramos que 110 (65%) mujeres postmenopáusicas tienen insomnio, su IC_{95%} es de 58 – 72% (marco rojo), lo que nos

Tibolone vs. estrogens

Measurement Site and Tibolone Dose Follow-Up Months	Tibolone Dose (mg)	No. of Comparisons (References)	MD (95%CI), Participants Random Effect Model	Heterogeneity	
				I ² (%)	p Value
Lumbar spine 12 m	2.5	3 [16,17,22]	-1.21 (-2.87 to 0.46), 232	67	0.05
	24 m	3 [16,17,22]	-0.58, (-3.77 to 2.60), 232	88	0.0002
Tibolone vs. Non-Active Controls					
Lumbar spine 12 m	2.5	4 [14,16,21,24]	2.75 (2.22 to 3.29), 328	0	0.60
	1.25	2 [21,24]	3.21 (2.12 to 4.30), 169	34	0.22

MD: diferencia de medias

Figura 6. Ejemplo de presentación de resultados de intervalo de confianza para la diferencia de medias en un cuadro de resultados. Aquí se presentan los resultados de la diferencia de medias de la densidad mineral ósea en mujeres con tratamiento de tibolona comparadas con las de tratamiento de estrógenos en donde no hay diferencia entre los tratamientos (marco rojo), y de tibolona vs. controles sin tratamiento, mostrando una diferencia a favor de la tibolona (marco azul). Tomado de: Castrejón-Delgado (2021).¹⁷

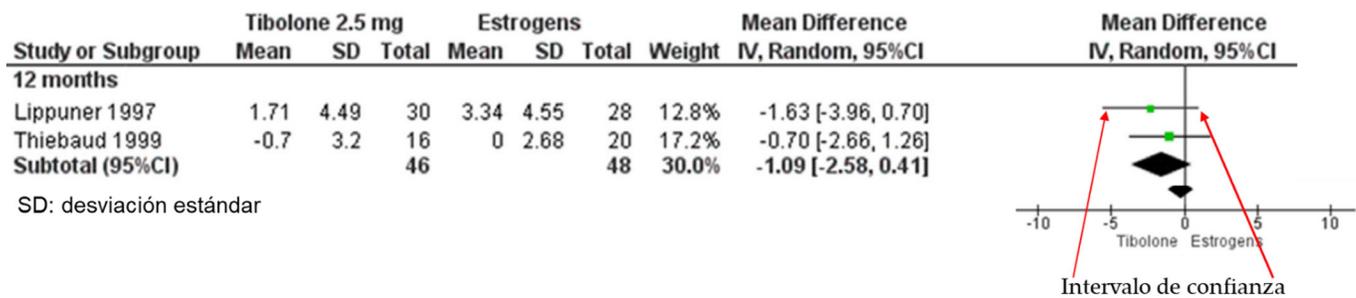


Figura 7. Ejemplo de presentación de resultados de intervalo de confianza para la diferencia de medias en una gráfica forest plot que muestra la densidad mineral ósea de mujeres con tratamiento de tibolona y de estrógenos en dos estudios diferentes y el resultado del metaanálisis. Los IC_{95%} muestran que no hay diferencia entre los tratamientos. Tomado de: Castrejón-Delgado, 2021.¹⁷

indica que del 58 al 72% de las mujeres postmenopáusicas presentarán insomnio. Aunque el intervalo es amplio, los valores tienden a valores altos y mayores del 50%, lo que nos indica que no son debidos al azar. No sucede lo mismo con la ingesta de alcohol, en donde, cualquiera de los dos grupos estudiados tiene una prevalencia entre 4 y 5%, con un IC_{95%} muy pequeño (marco verde), menor del 50% que podría indicar que esa prevalencia es al azar. En el caso de las proporciones, si el IC incluye el 50% o es menor de este valor, se considera una alta posibilidad de que el resultado sea debido al azar.

En el caso de los riesgos, los parámetros a calcular son el riesgo relativo (*risk ratio*, RR), la razón de momios (*odds ratio*, RM, OR) o la razón de daño (*hazard risk ratio*, HR). Todos se interpretan de la misma manera, considerando que

un factor es de riesgo si el resultado es >1 y protección si el valor es <1,^{18,19} aunque el cálculo, incluyendo su respectivo IC_{95%}, es diferente. En general, el parámetro de riesgo calculado (estimator puntual) indica el número de veces que es más probable que se produzca un efecto en un grupo expuesto a un posible factor de riesgo frente a un grupo no expuesto.^{19,20} En estos casos, el valor de 1 es considerado un riesgo nulo ya que refleja una razón entre expuestos y no expuestos 1:1 en el contexto de la causalidad,¹⁸ entonces, si el IC del riesgo calculado incluye al 1, no se puede hablar de riesgo,^{19,21} no importa que tan grande es el valor de riesgo calculado. También debe considerarse que el riesgo no es ni menor ni mayor al IC obtenido, y mientras más grande es el intervalo, mayor es la variación y probablemente el valor calculado no es riesgo. Así mismo, la amplitud del IC y el tamaño

Hot Flashes Intensity ^a	OR (95% confidence interval)		
	Mild (< 2) (n = 58)	Moderate (= 2) (n = 52)	Severe (≥ 3) (n = 60)
Oxidative stress (SS ≥ 4)	18 (31%)	18 (35%)	32 (53%)
Model ^b			
A	1.00	1.18 (0.53–2.61)	2.54 (1.20–5.39) ^c
B	1.00	1.46 (0.63–3.36)	4.67 (1.85–11.71) ^d
C	1.00	1.59 (0.67–3.80)	5.12 (1.99–13.17) ^e

^c $p < 0.05$ obtained with chi square for trend. ^d Significance into the model $p < 0.05$ using logistic regression.

^e Significance into the model $p < 0.05$ using logistic regression.

Figura 8. Ejemplo de la presentación de resultados de razón de momios con intervalo de confianza para la severidad de los bochornos como factor de riesgo para estrés oxidativo. Marco azul, intervalo de confianza que incluye al 1; marco rojo, intervalo de confianza que no incluye al 1. Tomado de: Sánchez-Rodríguez (2019).¹⁶

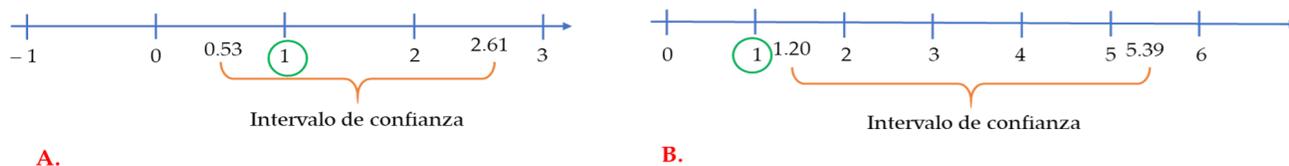


Figura 9. Representación de la recta numérica para identificar la localización de los intervalos de confianza. A. Ejemplo de intervalo de confianza que incluye al 1 (círculo verde). B. Ejemplo del intervalo de confianza que no incluye al 1.

muestral están relacionados con el valor de p , de tal manera que, si el IC es muy amplio y el tamaño de muestra no es suficiente, el valor de p será > 0.05 o no estadísticamente significativo la mayoría de las veces.

En la figura 8 se encuentra un ejemplo de lo señalado anteriormente. Son los resultados de diferentes análisis de riesgo para la severidad de los bochornos en mujeres postmenopáusicas y el estrés oxidativo. Los resultados presentados son de la razón de momios con su respectivo $IC_{95\%}$, y podemos observar que los bochornos moderados no pueden ser considerados como riesgo para estrés oxidativo en ninguno de los modelos construidos ya que los IC incluyen al 1 (marco azul), y los bochornos severos si son factor de riesgo ya que los IC no incluyen al 1 (marco rojo).

Analicemos ahora detalladamente el concepto de la inclusión del 1 en el IC tomando los datos del modelo A de la misma figura 8. Para los bochornos moderados, el $IC_{95\%}$ va de 0.53 – 2.61, si pensamos en la recta numérica, podemos observar que el 1 está dentro del intervalo (Figura 9A), por lo que los bochornos moderados no son factor de riesgo para estrés oxidativo. En el caso de los bochornos severos en el mismo modelo, el $IC_{95\%}$ es 1.20 – 5.39, este intervalo ya pasó el 1 (Figura 9B), por lo que el resultado de la RM indica que los bochornos severos son factor de riesgo para estrés oxidativo.

Como se mencionó anteriormente, en la discusión de la validez del valor de p como un indicador para la toma de decisiones clínicas, se propone la interpretación de este valor junto con el IC, siendo necesarios en el reporte de los resultados, por lo que las guías o normas para publicación de artículos como el STROBE (*Strengthening the Reporting of OBservational studies in Epidemiology*), guía para estudios observacionales, o el CONSORT (*Consolidated Standards of Reporting Trials*), guía para ensayos clínicos, anteponen el uso del IC a los valores p , ya que, como hemos visto, son una alternativa mucho más directa y fácil de entender, además de que se obtiene el efecto del tamaño muestral y la imprecisión de la medición;^{19,22} por lo que la recomendación cuando hacemos la lectura de un estudio de riesgos, principalmente, es revisar detalladamente el apartado de resultados en la búsqueda, no sólo del valor de riesgo calculado, sino el intervalo de confianza y el valor de p para hacer una correcta interpretación de ellos.

gthening the Reporting of OBservational studies in Epidemiology), guía para estudios observacionales, o el CONSORT (*Consolidated Standards of Reporting Trials*), guía para ensayos clínicos, anteponen el uso del IC a los valores p , ya que, como hemos visto, son una alternativa mucho más directa y fácil de entender, además de que se obtiene el efecto del tamaño muestral y la imprecisión de la medición;^{19,22} por lo que la recomendación cuando hacemos la lectura de un estudio de riesgos, principalmente, es revisar detalladamente el apartado de resultados en la búsqueda, no sólo del valor de riesgo calculado, sino el intervalo de confianza y el valor de p para hacer una correcta interpretación de ellos.

AGRADECIMIENTOS

El manuscrito fue revisado y editado en el Programa para la Investigación Bibliográfica Científica sobre Salud (PIBCIS) de la FES Zaragoza, UNAM. El artículo es parte del material de apoyo didáctico elaborado para el proyecto PAPIME PE203421.

REFERENCIAS

1. López-Hernández D, Torres-Fonseca A, Brito-Aranda L, López-Hernández ML. Cómo redactar y organizar un artículo científico original. *Rev Esp Med Quir.* 2014; 19: 236-243.
2. Henríquez-Fierro E, Zepeda-González MI. Elaboración de un artículo científico de investigación. *Cienc Enf.* 2004; 10(1): 17-21.
3. Abreu JL. Investigación: resultados, discusión & conclusiones. *Int J Good Consc.* 2012; 7(3): 131-138.
4. Cobos-Valeri E, González-Alastrué JA, Muñoz-Gracia P, Bigorra-Llosas J, et al. *Bioesta-*



- dística para no estadísticos. Barcelona: Mas-son; 2007. p. 157-192.
5. Elorsa H. Estadística para las ciencias socia-les y del comportamiento. 2a ed. México: Oxford University Press; 2000. p. 296-317
6. Guller U, DeLong ER. Interpreting statistics in medical literature: A vade mecum for surgeons. *J Am Coll Surg*. 2004; 198(3): 441-458.
7. García-García JJ. Significancia estadística y significancia clínica. En: Moreno-Altamirano L. *Epidemiología clínica*. 3ª Ed. México: McGraw-Hill; 2013. p. 283-291.
8. Armitage P, Berry G, Matthews JNS. *Statistical methods in medical research*. 4 th Ed. Mas-sachusetts: Blackwell Science; 2002. p. 83-92.
9. Browner WS, Newman TB, Cummings SR, Hulley SB. Getting ready to estimate sample size: hypothesis and underlying principles. In: Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. *Designing clinical research*. 3rd Ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2007. p. 51-63.
10. Loria A. Estadística mínima LXII. Las P de las pruebas estadísticas. *Lab-acta*. 2004; 16: 41-44.
11. Sánchez-Rodríguez MA, Castrejón-Delgado L, Zacarías-Flores M, Arronte-Rosales A, Men-doza-Núñez VM. Quality of life among post-me-nopausal women due to oxidative stress boost-ed by dysthymia and anxiety. *BMC Women's Health*. 2017; 17: 1. doi: 10.1186/s12905-016-0358-7
12. Sánchez-Rodríguez MA, Ruiz-Ramos M, Co-rrera-Muñoz E, Mendoza-Núñez VM. Oxidative stress as a risk factor for osteoporosis in elderly Mexicans as characterized by antioxidant enzy-mes. *BMC Musculoskeletal Disorders*. 2007; 8: 124. doi: 10.1186/1471-2474-8-124.
13. Ruiz-Rodríguez AK. Tiempo transcurrido para la disminución del estrés oxidativo asocia-do a la intensidad de los bochornos con el uso de la tibolona en la posmenopausia. Tesis para el grado de Maestría en Ciencias de la Salud. Facultad de Medicina, UNAM; 2018. p. 62.
14. Castañeda JA, Fabián-Gil J. Una mirada a los intervalos de confianza en investigación. *Rev Colomb Psiquiatr*. 2004; 33(2): 193-201.
15. Aedo MS. Intervalos de confianza para propor-ciones. *Rev Obstet Ginecol*. 2008; 3(1): 68-70.
16. Sánchez-Rodríguez MA, Zacarías-Flores M, Arronte-Rosales A, Mendoza-Núñez VM. As-sociation between hot flashes severity and oxi-dative stress among Mexican postmenopausal women: A cross-sectional study. *PLoS ONE*. 2019; 14(9): e0214264. doi: 10.1371/journal.pone.0214264
17. Castrejón-Delgado L, Castelán-Martínez OD, Clark P, Garduño-Espinosa J, Mendoza-Núñez VM, Sánchez-Rodríguez MA. Effect of tibolone on bone mineral density in postmeno-pausal women: systematic review and meta-analysis. *Biology*. 2021; 10: 211. doi: 10.3390/biology10030211.
18. Vargas-Voráckova F. Concepto de causali-dad en medicina. En: Moreno-Altamirano L. *Epi-demiología clínica*. 3ª Ed. México: McGraw-Hill; 2013. p. 153-165.
19. Talavera JO, Rivas-Ruiz R. Investigación clí-nica VI. Relevancia clínica. *Rev Med Inst Mex Seguro Soc* 2011; 49 (6): 631-635.
20. Martínez-González MA, Seguí-Gómez M, Delgado-Rodríguez M. ¿Cómo mejorar los in-tervalos de confianza? *Med Clin (Barc)*. 2010; 135(1): 30-34.
21. Loria A. Estadística mínima XLVII. El inter-valo de confianza de la razón de momios. *Lab-acta*. 2000; 12: 79-83.
22. González-Marrón A, Real J, Forné C, Ro-so-Llorach A, Navarrete-Muñoz EM, Martínez-Sánchez JM. Confidence Interval reporting for measures of association in multivariable regres-sion models in observational studies. *Med Clin (Barc)*. 2019; 153(6): 239-242. doi: 10.1016/j.medcli.2018.06.018.