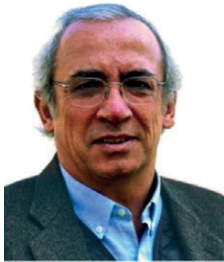




COLABORACIÓN  
ESPECIAL



**Prof. José Carlos de la Macorra García**  
Licenciado en Medicina,  
Especialista en Estomatología,  
Especialista en Medicina del  
Trabajo, Doctor en Medicina y  
Cirugía (1988) en la Universidad  
Complutense de Madrid.  
Profesor Emérito (Universidad  
Complutense de Madrid).

**Correspondencia:**  
Prof. José Carlos de la Macorra García  
Departamento de Odontología  
Conservadora y Prótesis.  
Facultad de Odontología.  
Plaza Ramón y Cajal s/n  
Ciudad Universitaria. 28040 Madrid.  
macorra@ucm.es

## La significación insignificante<sup>1</sup>

De la Macorra García JC.  
La significación insignificante.  
Cient. Dent. 2025; 22; 1; 11-14

Es difícil transcribir los resultados de la investigación a la clínica, al paciente. En inglés este proceso se denomina *from bench to bedside*, dando origen al importante campo de la investigación traslacional<sup>2</sup>, un proceso que afecta todos los niveles de la investigación, desde los descubrimientos científicos básicos hasta sus posibles impactos en la salud pública.

Pero hay, además, una dificultad adicional básica, previa, que no se contempla. Debido a que la metodología de la investigación es un proceso complejo y frecuentemente poco intuitivo, hay una dificultad en la interpretación directa, inmediata, de los resultados de los estudios: como se ha dicho, entender lo que significa la investigación es frecuentemente peliagudo<sup>3</sup>. Esto da pie a varias situaciones confusas, de las que repasaremos algunas.

### LA P TIENE UN PRECIO

Ya sabemos que la famosa *p* es, no más, no menos, la probabilidad de que el azar haya causado los resultados de nuestro experimento o nuestra observación<sup>3</sup>. Nos da así un criterio sobre la credibilidad de los resultados, y de la hipótesis que los predijo.

Esto se hace mediante unos cálculos estadísticos más o menos complejos que tienen en cuenta, esencialmente, dos cosas: cuán bien hemos medido (*cuánto se dispersan los datos, generalmente expresado con la desviación estándar o*

*una medida asimilable*), y cuántos casos hemos estudiado.

Ya hemos mencionado cómo la variabilidad en los datos afecta a la *p* y cómo el mejorarla (disminuirla) cuesta dinero<sup>4</sup>. Veremos ahora cómo el número de casos que se ha estudiado interviene en ello, de manera que es posible aumentar la significación simplemente aumentando el número de casos<sup>5</sup>.

Utilizaremos para ello unos datos generados al azar, simulando un estudio comparando un tratamiento (A), bien conocido y usado desde hace tiempo, y uno (B), novedoso, ambos indicados contra el dolor. Se mide el tiempo, en horas, durante el que los pacientes no sienten dolor. En un primer estudio, encargado por la empresa fabricante de B, se comparan ambos medicamentos sobre 5 pacientes por grupo.

Aunque estrictamente esta sería una prueba de superioridad, que requeriría unos cálculos algo más complejos<sup>6,7</sup>, plantearemos una comparación básica entre los grupos mediante pruebas *t*. Los resultados quedan reflejados en la Tabla 1.

Es decir, el medicamento B aumenta de media el tiempo sin dolor 1.6 hs (22.3 hs el medicamento B respecto a 20.7 hs el A) pero la probabilidad de que estos resultados sean debidos al azar es muy alta (0.97); este resultado casi seguro es debido a que solo se han comparado los resultados en 10 pacientes, 5 por grupo. Sin embargo, son promisorios: el medi-

1. La denominación tan sugestiva de *significación insignificante* está tomada del magnífico trabajo de Skrabanek y McCormick (1), donde se presenta como una falacia.

Tabla 1. Comparación básica entre grupos. Estudio 1.

Estudio 1					
Medicamento	n	m(hs)	ds(hs)	diferencia(hs)	p
A	5	20.7	1.4	1.6	0.97
B	5	22.3	1.1		

n: número de casos; m: media de tiempo sin dolor, en horas; ds: desviación estándar, en horas, p: significación

El medicamento B parece funcionar mejor que el A, aunque los resultados no sean, como decimos, estadísticamente significativos.

Como dueño de la empresa que fabrica el medicamento B, encargo entonces un segundo estudio con la misma metodología, pero esta vez con 25 pacientes por grupo -es decir, un estudio algo más caro que el primero-. Los resultados son los reflejados en la Tabla 2.

Tabla 2. Comparación entre grupos. Estudio 2

Estudio 2					
Medicamento	n	m(hs)	ds(hs)	diferencia(hs)	p
A	25	20.2	2,6	1.3	0.4
B	25	21.5	2,5		

n: número de casos; m: media de tiempo sin dolor, en horas; ds: desviación estándar, en horas, p: significación

El medicamento B sigue siendo mejor (la diferencia del tiempo medio sin dolor esta vez es de 1.3hs, siendo más corto el A), y la p está algo por debajo de la significación estadística ( $p=0.4$ ). Es decir: es suficientemente poco probable que estos resultados sean debidos al azar, por lo que entendemos que se debe a la diferencia entre los medicamentos. La explicación es que, como hemos usado 50 pacientes en total, es difícil que con tantos pacientes el azar intervenga en exceso. Obsérvese que las desviaciones estándar (ds) de ambos grupos son, en este segundo estudio, mayores que en el primero. Aun así, la significación es mayor.

Pero, como dueño de la empresa interesada en comercializar el medicamento B, quiero que mi significación sea aún menor, no quiero que nadie me diga que estos resultados pueden ser ni remotamente debidos al azar. Para ello, encargo otro estudio, esta vez con 500 pacientes por grupo -nótese, mucho más caro de llevar a cabo que los dos primeros-, cuyos resultados son los analizados en la Tabla 3.

Como en los estudios anteriores, el medicamento B tiene mejor comportamiento que el A (una mejoría de 1.02 hs),

Tabla 3.- Comparación entre grupos. Estudio 3

Estudio 3					
Medicamento	n	m(hs)	ds(hs)	diferencia(hs)	p
A	500	20.1	2,5	1.02	0.0000000001
B	500	21.1	2,5		

n: número de casos; m: media de tiempo sin dolor, en horas; ds: desviación estándar, en horas, p: significación

esta vez con una p muy significativa. Es altamente improbable que este resultado sea debido al azar y podemos concluir que el medicamento B tiene mejor comportamiento que el A.

Puede verse, tras observar los desenlaces de estos tres estudios, que la p ha ido descendiendo hacia la significación a medida que ha ido aumentando el número de casos. Este descenso no es debido a cambios notables en la desviación estándar, que en todo caso ha ido haciéndose ligeramente mayor con cada estudio.

En una representación imaginaria ya utilizada en esta serie de artículos<sup>4</sup>, supongamos un duende en nuestro ordenador que está interpretando los datos de los tres estudios anteriores. Se dirá a sí mismo: *de acuerdo, en el tercer estudio las diferencias entre los grupos son realmente pequeñas (1.02 hs de mejoría), menores que en los dos estudios anteriores, pero esta diferencia se ha encontrado utilizando un gran número de pacientes, y la desviación estándar es relativamente aceptable (2.5 hs). Estos resultados casi seguramente se deben a que hay diferencias reales entre los dos medicamentos: asignaré un p muy baja.*

## LA NO SIGNIFICACIÓN NO ES LA NO EXISTENCIA

No es raro asistir a presentaciones científicas o conferencias en las que, al presentarse unos resultados de investigación que no son estadísticamente significativos se traduce casi como que son nulos. Esto es inapropiado, pues la no significación solamente quiere decir que los resultados pueden -probablemente- haber sido causados por el azar, no que no existan.

En el primer estudio mostrado en la Tabla 1 hay una mejora del medicamento B respecto a A (1,6 hs. menos de dolor de media). El efecto, la mejora, existe. Y el hecho de que no sea estadísticamente significativa no la hace desaparecer.

En términos estadísticos, la no significación no prueba la hipótesis nula (*la hipótesis de que no hay diferencias*)<sup>8</sup>. Solo la hace más creíble, en ese contexto: ya se ha visto

que, simplemente aumentando el número de casos, dicha hipótesis se vuelve increíble.

Esta es una de las razones por las que la significación consensuada ( $p < 0.05$ ) es actualmente tan contestada<sup>5,9,10</sup> y por qué se ha extendido la exigencia del cálculo previo del tamaño muestral<sup>11</sup> y la interpretación de los resultados mediante el uso de los intervalos de confianza<sup>12</sup>.

## LA SIGNIFICACIÓN INSIGNIFICANTE: LO QUE IMPORTA ES LA CLÍNICA

En los ejemplos anteriores se ve cómo el efecto (*la mejora del tiempo sin dolor que buscamos con el medicamento B*) disminuye con cada estudio: pasa de 1,6 a 1,3 a 1,02 hs., y ya hemos visto que, pese a ello, la significación ha ido dis-

minuyendo: no es el tamaño *-la importancia, la relevancia clínica-* del efecto per se lo que hace que la significación sea mayor o menor.

Debe observarse que esto no tiene porqué ocurrir de esta manera en todos los casos: el tamaño del efecto podría ser menor en el primer estudio o mayor en el último. La ordenación actual de mayor a menor es debida al azar. Lo que es relevante es que el efecto puede ser menor, clínicamente menos importante, con una significación menor.

Este es uno de los conceptos clave: la significación estadística no se traduce necesariamente en relevancia clínica. Un efecto puede ser muy pequeño, incluso clínicamente irrelevante, pero su estudio puede dar unos resultados estadísticamente muy significativos, siempre que cuente con el suficiente número de casos.



## BIBLIOGRAFÍA

1. Skrabanek P, McCormick J. Follies and fallacies in medicine. 3 ed. Eastbourne: Tarragon Press,1998.
2. Drolet BC, Lorenzi NM. Translational research: understanding the continuum from bench to bedside. *Transl Res.* 2011;157(1):1-5. <https://doi.org/10.1016/j.trsl.2010.10.002>.
3. De la Macorra García JC. La ciencia, el método científico, el azar y la navaja de Ockham. *Cient Dent.* 2024;21(2):90-92.
4. De la Macorra García JC. Medir las cosas. La precisión y la exactitud. *Cient Dent.* 2024;21(3):160-163.
5. Demidenko E. The p-value you can't buy. *Am Stat.* 2016;70(1):33-38. <https://doi.org/10.1080/00031305.2015.1069760>.
6. Wang B, Wang H, Tu XM, Feng C. Comparisons of superiority, non-inferiority, and equivalence trials. *Shanghai Arch Psychiatry.* 2017;29(6):385-388. <https://doi.org/10.11919/j.issn.1002-0829.217163>.
7. Goeman JJ, Solari A, Stijnen T. Three-sided hypothesis testing: Simultaneous testing of superiority, equivalence and inferiority. *Statistics in Medicine.* 2010;29(20):2117-2125. <https://doi.org/10.1002/sim.4002>.
8. Amrhein V, Greenland S, McShane B. Retire statistical significance. *Nature.* 2019;567(7748):305-307. <https://doi.org/10.1038/d41586-019-00857-9>.
9. McShane BB, Gal D, Gelman A, Robert C, Tackett JL. Abandon statistical significance. *Am Stat.* 2019;73(sup1):235-245. <https://doi.org/10.1080/00031305.2018.1527253>.
10. Hurlbert SH, Levine RA, Utts J. Coup de grâce for a tough old bull: "statistically significant" expires. *Am Stat.* 2019;73(sup1):352-357. <https://doi.org/10.1080/00031305.2018.1543616>.
11. Serdar CC, Cihan M, Yücel D, Serdar MA. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. *Biochem Med.* 2021;31(1):010502. <https://doi.org/10.11613/bm.2021.010502>.
12. Anderson AA. Assessing statistical results: magnitude, precision, and model uncertainty. *Am Stat.* 2019;73(sup1):118-121. <https://doi.org/10.1080/00031305.2018.1537889>.