



## Editorial

## ¿Estadísticamente significativo o clínicamente relevante?

### Statistically Significant or Clinically Important?



Este editorial es la tercera y última entrega de la serie de 3 editoriales. En la primera se explicó el ciclo del método científico para contextualizar, en él, el papel de la estadística, junto con los conceptos de precisión y validez interna y externa<sup>1</sup>. En la segunda se intentó dejar claro que si una diferencia no es estadísticamente significativa, esto no es sinónimo de equivalencia y no debe considerarse que ambos tratamientos «son iguales o equivalentes», en el contexto por ejemplo de un ensayo clínico en el que se compare un nuevo tratamiento con el tratamiento convencional<sup>2</sup>.

Este último editorial pretende mostrar la diferencia entre estadísticamente significativo y clínicamente relevante<sup>3</sup>.

Pongamos como ejemplo una «variable respuesta principal» o «primary endpoint», que en este caso es el cambio desde la visita basal hasta la visita final en el volumen espiratorio forzado en el primer segundo (FEV1)<sup>4</sup>, por ejemplo, en el contexto comparativo de 2 tratamientos en asma.

Planteando el diseño del ensayo clínico como un ensayo clínico controlado y aleatorizado de 2 brazos paralelos, obtendremos como medida descriptiva la media de la puntuación en el FEV1 en el grupo de intervención y la compararemos con la media en el grupo control, obteniendo la «diferencia de medias entre grupos» en litros o mililitros (mL), que es una variable cuantitativa continua al ser susceptible de tomar valores decimales. En cuanto a test estadísticos, el test t de Student-Fisher es el procedimiento más habitual para comparar ambas medias.

En la [figura 1](#), el umbral para la diferencia mínima clínicamente relevante (+100 mL) se muestra con barras en color negro resaltado, y el umbral de la hipótesis nula clásica de diferencia de mL = 0 se muestra con barras y puntos (resaltado en rojo en la versión electrónica del editorial). Cada intervalo de confianza (IC) representa los límites inferior y superior de la verdadera diferencia de medias poblacional de respuesta a los tratamientos (con un 95% de confianza). En los cuatro IC95%, el límite inferior de cada intervalo es mayor que cero. Si dividiéramos la diferencia de medias entre su error estándar, que recordemos que cuantifica el papel del azar en nuestros resultados, comprobaríamos cómo el resultado sería un estadístico «t de Student» mayor que 1,96; con una  $p < 0,05$ . Así pues, los cuatro IC95% son estadísticamente significativos.

Que un resultado sea estadísticamente significativo, únicamente va a querer decir que el error estándar es pequeño en comparación con la diferencia de medias. Si el error estándar es lo suficientemente pequeño, nuestros resultados serán siempre estadísticamente significativos, pero ello no implica que

necesariamente sean clínicamente relevantes. En este caso, como hemos mencionado, se considera que una diferencia en el FEV1 tiene que ser de al menos 100 mL para que tenga un impacto en la calidad de vida de los pacientes.

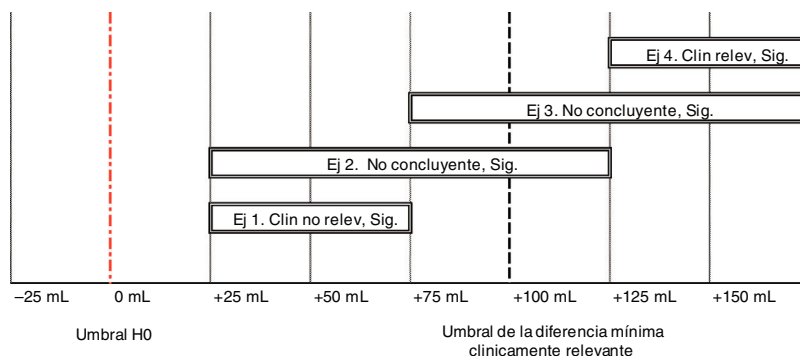
Siguiendo con el ejemplo de la [figura 1](#), veremos cómo a pesar de que todos los IC95% son estadísticamente significativos, no todos son concluyentes en cuanto a la relevancia clínica del tratamiento. La relevancia crítica se interpreta en base al tamaño del efecto o «effect size» (nuestra diferencia de medias), usando los límites de los IC95% en relación con el umbral delta que se corresponde con la diferencia mínima clínicamente relevante, conocida en inglés como «minimal clinically important difference» (MCID)<sup>5,6</sup>.

El primer IC95% (ejemplo 1) se corresponde con una diferencia de +50 mL, con un límite inferior y uno superior de +25 a +75 mL respectivamente (teniendo en cuenta el papel del azar en nuestros resultados con un 95% de confianza). Es decir, el nuevo tratamiento sería mejor que el tratamiento convencional, pero no tanto como para considerarlo clínicamente relevante, porque su límite superior (+75 mL) es menor que la diferencia mínima clínicamente relevante de +100 mL. Este es el ejemplo paradigmático de cómo algo puede ser estadísticamente significativo, pero clínicamente no relevante.

Recordemos que el error estándar se puede disminuir reduciendo la variabilidad o aumentando el tamaño muestral, luego cualquier diferencia de medias que no sea cero se podrá volver estadísticamente significativa. Una diferencia de medias totalmente irrelevante de 0,05 mL será estadísticamente significativa si, por ejemplo, logramos que su error estándar asociado sea de 0,025 mL.

Únicamente un IC95% (el del ejemplo 4), además de estadísticamente significativo, es clínicamente relevante de forma concluyente porque su límite inferior es mayor que 100 mL. En el ejemplo 2 la mayor parte del intervalo concuerda con la hipótesis de que la diferencia no es clínicamente relevante. En el ejemplo 3 la mayor parte del intervalo concuerda con que el efecto sí es clínicamente relevante. No obstante, al cruzar sendos IC95% el umbral para la diferencia mínima clínicamente relevante, ambos ejemplos serían clínicamente no concluyentes.

Esto mismo se aplica a las medidas de asociación. La relevancia clínica en este caso ha de tenerse siempre en cuenta ante resultados cercanos a 1 en medidas de asociación, y especialmente si se trata de *odds ratios*, ya que recordemos que esta medida puede sobreestimar la magnitud de la asociación con respecto a otras como la razón de riesgos<sup>7</sup>.



**Figura 1.** Interpretación de la relevancia clínica de ejemplos de intervalos de confianza al 95%, en el que se ha establecido el umbral para la diferencia mínima clínicamente relevante en una diferencia de medias en el FEV1 mayor que +100 mL. Sig. = estadísticamente significativo ( $p < 0,05$ ).

Así pues, a modo de conclusión, cualquier diferencia de medias que no sea cero, o cualquier medida de asociación que no sea 1, se podrá volver estadísticamente significativa aumentando el tamaño muestral o disminuyendo la variabilidad de los datos. Que un resultado sea estadísticamente significativo no es sinónimo de relevancia clínica, pues no todos los resultados estadísticamente significativos serán clínicamente relevantes, cuya interpretación debe hacerse en base a un umbral específico de relevancia clínica. El problema y las diferentes metodologías a la hora de establecer un umbral para la diferencia mínima clínicamente relevante es motivo de interés científico<sup>8,9</sup> y excede las pretensiones de este editorial. Por su utilidad práctica en el caso de la investigación aplicada a enfermedades respiratorias, conviene recordar los cuestionarios de control de síntomas como por ejemplo el *Asthma Control Test* (ACT, por sus siglas en inglés). Estas herramientas validadas reportan, entre sus características métricas, la llamada «sensibilidad al cambio» o «responsiveness» en inglés<sup>10</sup>, estableciendo asimismo la diferencia mínima clínicamente relevante que en este caso es de 3 puntos<sup>11</sup>.

## Bibliografía

- Santibáñez M, García-Rivero JL, Barreiro E. No se debe empezar la casa por el tejado (si queremos publicar en una revista de impacto). *Arch Bronconeumol*. 2019. <http://dx.doi.org/10.1016/j.arbres.2019.05.019>.
- Santibáñez M, García-Rivero JL, Barreiro E.  $p$  de significación: ¿mejor no usarla si se interpreta mal? *Arch Bronconeumol*. 2019. <http://dx.doi.org/10.1016/j.arbres.2019.11.003>.
- Chan KB, Man-Son-Hing M, Molnar FJ, Laupacis A. How well is the clinical importance of study results reported? An assessment of randomized controlled trials. *CMAJ*. 2001;165:1197–202.
- Puente Maestú L, García de Pedro J. Lung function tests in clinical decision-making. *Arch Bronconeumol*. 2012;48:161–9.
- Kieser M, Hauschke D. Assessment of clinical relevance by considering point estimates and associated confidence intervals. *Pharm Stat*. 2005;4:101–7.
- Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10:407–15.
- Viera AJ. Odds ratios and risk ratios: what's the difference and why does it matter? *South Med J*. 2008;101:730–4.
- Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc*. 2002;77:371–83.
- Sloan JA. Assessing the minimally clinically significant difference: scientific considerations, challenges and solutions. *COPD*. 2005;2:57–62.
- Schatz M, Sorkness CA, Li JT, Marcus P, Murray JJ, Nathan RA, et al. Asthma Control Test: reliability, validity, and responsiveness in patients not previously followed by asthma specialists. *J Allergy Clin Immunol*. 2006;117:549–56.
- Schatz M, Kosinski M, Yarlas AS, Hanlon J, Watson ME, Jhingran P. The minimally important difference of the Asthma Control Test. *J Allergy Clin Immunol*. 2009;124:719–23.

Miguel Santibáñez<sup>a,\*</sup>, Juan Luis García-Rivero<sup>b</sup> y Esther Barreiro<sup>c,d,e</sup>

<sup>a</sup> Grupo de Investigación de Salud Global, Universidad de Cantabria, Instituto de Investigación Marqués de Valdecilla (IDIVAL), Santander, Cantabria, España

<sup>b</sup> Servicio de Neumología, Hospital de Laredo, Laredo, Cantabria, España

<sup>c</sup> Servicio de Neumología-Debilidad muscular y caquexia en las enfermedades respiratorias crónicas y el cáncer de pulmón, IMIM-Hospital del Mar, Barcelona, España

<sup>d</sup> Departament de Ciències Experimentals i de la Salut (CEXS), Universitat Pompeu Fabra (UPF), Parc de Recerca Biomèdica de Barcelona (PRBB), Barcelona, España

<sup>e</sup> Centro de Investigación en Red de Enfermedades Respiratorias (CIBERES), Instituto de Salud Carlos III (ISCIII), Barcelona, España

\* Autor para correspondencia.

Correo electrónico: [santibanezm@unican.es](mailto:santibanezm@unican.es) (M. Santibáñez).