

4.- Uso e interpretación de los intervalos de confianza

Use and interpretation of confidence intervals

José Antonio González¹

jose.a.gonzalez@upc.edu

1 Departamento de Estadística e Investigación Operativa, Barcelona-Tech, UPC

Explicamos la importancia de acompañar los resultados con intervalos de confianza, que bien podríamos rebautizar como *márgenes de incertidumbre*, en negativo; o *de compatibilidad*, en positivo.

Lo aconsejan las guías de publicación

La guía CONSORT¹, en su punto 17 pide comunicar el tamaño del efecto («effect size») junto con alguna medida de incertidumbre, como el intervalo de confianza del 95% (IC_{95%}). El tamaño del efecto mide cuánto mejor evolucionan los pacientes del grupo tratado respecto al grupo de referencia. El intervalo de confianza refleja la incertidumbre originada por el proceso de asignación al azar o aleatorización. En un ensayo clínico, la asignación al azar de las intervenciones en comparación genera unos resultados que habrían sido diferentes si la asignación de las intervenciones a los casos hubiera sido distinta. Así, si los resultados del estudio son irrepetibles, ¿cómo valorar su replicabilidad? Recordemos que la replicabilidad es fundamental en la ciencia. Y por extensión en la clínica, donde lo que no es repetible no tiene valor para los futuros pacientes.

Los intervalos proporcionan los valores del efecto poblacional compatibles con los resultados del estudio

La teoría estadística clásica distingue entre un efecto «poblacional», en todos los casos con los mismos criterios de elegibilidad; y un efecto en los datos observados, en la muestra obtenida. El efecto en toda la población es único, es una constante (un *parámetro*, según la terminología estadística). En cambio, el efecto observado será distinto para cada asignación al azar, será variable: en cada aleatorización será distinto. No es, por tanto, un valor repetible; es tan solo el resultado observado. Como analogía, podríamos decir

que el efecto poblacional es una verdad escondida, pero los efectos observables en muestras son mentiras a la vista. El intervalo de confianza refleja aquellos valores del parámetro probabilísticamente compatibles con el valor observado.

Pongamos cierto ensayo clínico para valorar el descenso de la mortalidad de un nuevo tratamiento en una cierta enfermedad. El análisis principal es el cociente, entre los 2 grupos de tratamiento, de las probabilidades de morir durante el seguimiento: lo que conocemos como «riesgo relativo». Los resultados estiman puntualmente que este cociente vale 0,61: el nuevo tratamiento multiplica por 0,61 la mortalidad del grupo de referencia. Por tanto, la baja un 39% (= 100-61%). Y proporcionan un IC95% que va desde 0,33 hasta 1,11: los resultados observados son compatibles con que la nueva intervención multiplica la probabilidad de morir entre 0,33 y 1,11. Si pudiéramos aplicar el nuevo tratamiento a toda la población, creemos que su mortalidad quedaría entre 0,33 y 1,11 veces la mortalidad que hubiéramos observado si hubiéramos administrado a toda la población el tratamiento de referencia. Los resultados son compatibles tanto con que la mortalidad baja a la tercera parte (0,33), como con una subida de un 11% (1,11).

Recordemos brevemente que incrementar el número de participantes supone aumentar la información disponible, lo que se refleja en un intervalo de confianza más estrecho, más preciso. En cambio, si aumenta la dispersión entre los participantes (la variabilidad), el intervalo se hace más amplio, reflejando mayor incertidumbre.

¿Es informativo el intervalo?

Debemos valorar si el intervalo aporta información, nuevo conocimiento. O si, por el contrario, el resultado proporcionado no nos aporta nada.

Supongamos que estamos en un primer estudio para valorar si la nueva intervención aumenta la probabilidad de curarse. El intervalo del cociente de probabilidades va desde 0.2 hasta 10. En su extremo superior, diríamos que los resultados del estudio son compatibles con que la nueva intervención multiplica por 10 las probabilidades de sanar. Y en su extremo inferior, que

las divide por 5 ($0.2=1/5$). Los resultados observados son compatibles con que la intervención nueva tanto podría beneficiar a los pacientes como perjudicarlos. Dos extremos opuestos: podríamos interpretar que la lección aprendida es nula; y etiquetar el estudio como anodino, poco informativo.

Pongamos otro ejemplo: el $IC_{95\%}$ del cociente de las probabilidades de curar va desde 20 a 30. Parece razonable decir que sí, que algo hemos aprendido, ya que el efecto observado sería relevante, tanto si fuera tan grande como 30, como si fuera tan grande como 20. Sin embargo, si un estudio previo había establecido que esa intervención podía subir la probabilidad de curar entre 22 y 26 veces, añadir ahora que podría subirla entre 20 y 30 veces parece poca aportación. O quizás todo lo contrario, ya que nos está recordando lo más importante en Ciencia: han reproducido los resultados previos.

Nótese que en ambos casos al final debe existir un juicio de valor que, en un hipotético manuscrito, el propio investigador debería sustentar en la discusión.

Los intervalos reflejan la incertidumbre aleatoria.

El documento explicativo de Strobe³ en su ítem 10 resume que el $IC_{95\%}$ aporta al lector la precisión estadística. Por «precisión estadística», se refiere a la incertidumbre introducida por los métodos de muestreo. Y a continuación se añade que todos los estudios pueden tener incertidumbres adicionales que deben, por lo menos, comentarse en la discusión. Además, en el ítem 20 de la discusión añade que la incertidumbre global siempre será mayor.

En resumen

Interprete los intervalos de confianza como informativos de: (1) los valores poblacionales compatibles con los resultados observados; y (2) la magnitud de la incertidumbre restante tras el estudio. Y si no existe un proceso de azar en la generación de los datos, o si el estudio tiene alguna impureza, admita cautamente las limitaciones presentes en el estudio y recuerde que la incertidumbre sería mayor que la reflejada por el intervalo.

Referencias

- ¹ Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332. PMID: 20332509
- ² Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *PLoS Med*. 2009;6(7):e1000100. PMID: 19621070
- ³ Vandembroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *Epidemiology*. 2007;18(6):805-35.