

Una pincelada estadística con repercusiones extrametodológicas

A statistical brushstroke with extra-methodological repercussions

Silva Ayçaguer, Luis Carlos¹

¹Investigador del Centro Nacional de Información de Ciencias Médicas. Profesor titular de la Escuela Nacional de Salud Pública de La Habana, Cuba. Editor adjunto de la Revista Cubana de Salud Pública. icsilva@infomed.sld.cu

Sr. Director:

Me permito comenzar destacando el importante papel que está desempeñando la revista *Salud Colectiva*, por lo cual extiendo mi felicitación a todo el colectivo de trabajo que la hace posible. En particular, he leído con mucho interés el volumen 7 número 2 de 2011, destinado en buena medida al crucial tema de los ensayos clínicos. Uno de sus artículos, titulado "Cuatro palabras sobre ensayos clínicos: ciencia/negocio, riesgo/beneficio" (1), a cargo de los colegas Antonio Ugalde y Núria Homedes, me llamó especialmente la atención. Considero que allí se hace una trascendente contribución al examen de los manejos que realiza la industria, los comités de ética y los órganos reguladores.

En el trabajo, sin embargo, se suscribe una afirmación tomada de un documento de 2009 de la *Food and Drug Administration* (FDA) de EE.UU., con la cual no estoy de acuerdo. La cita en cuestión reza así:

Una limitación es el tamaño de la muestra de los ensayos. Por razones de economía y dificultades para reclutar pacientes que cumplan con los criterios de inclusión, un ensayo fase 3 raramente incluye más de 4.000 o 5.000 pacientes; la excepción son las vacunas que se ensayan en muestras mucho más grandes de personas sanas. Una muestra de 4.000 o 5.000 pacientes no es suficiente para representar la variedad genética, sociodemográfica (edad, sexo, condiciones sanitarias, etc.) y las condiciones de salud (enfermedades concomitantes, estado nutricional, etc.) de la población. (1 p.136)

Los juicios según los cuales "el tamaño de muestra es insuficiente" siguen siendo un criterio valorativo muy consolidado en comités de ética, agencias de evaluación de proyectos de investigación y árbitros o editores de revistas. La legitimidad de este tipo de pronunciamientos es, sin embargo, muy discutible.

Tales juicios son conflictivos y casi siempre improcedentes por al menos dos razones. En primer lugar, porque las fórmulas para determinar el tamaño muestral son intrínsecamente especulativas y comportan una inevitable carga de subjetividad (2). En segundo lugar, porque la noción de "tamaño suficientemente grande" solo tendría sentido en el supuesto de que pudieran sacarse reglas operativas o conclusiones definitivas de cada estudio aislado. Sin embargo, esta es una ilusión tan arraigada como incorrecta, por la simple razón de que la ciencia no funciona así.

Nuestras convicciones científicas pueden ser más o ser menos firmes, pero siempre son provisionales, y nuestras representaciones de la realidad tienen en cada momento un cierto grado de credibilidad, pero están abiertas a cambios y perfeccionamientos en la medida en que nuevos datos lo aconsejen. La consolidación del nuevo conocimiento es gradual y cualquier aporte metodológicamente riguroso es bienvenido. Unos serán más trascendentes y otros menos. Algunos se concentrarán en ciertas franjas de edad, o se realizarán en determinadas condiciones sociosanitarias, etc., y otros en condiciones diferentes. Pero todos pueden hacer alguna contribución en el cambiante proceso de construcción del consenso, independientemente del tamaño de muestra. En esta materia se avanza por lo general a través del metaanálisis (3) o por medio del enfoque

bayesiano (4), por mencionar los dos enfoques alternativos más conocidos.

Este elemento es relevante, ya que el tamaño de muestra mínimo que exigen los cánones ritualizados –obviamente, la FDA es uno de los órganos que los siguen, aunque más recientemente han matizado su posición (5)– es aquel *suficientemente grande como para "detectar" diferencias*; expresión ambigua que, en realidad, alude a un tamaño suficientemente grande como para declarar que la diferencia que mide el efecto es estadísticamente significativa.

Aparte de la falacia implícita en la demanda de hacer ensayos enormes, que reposa en una idea que ignora que el conocimiento se construye de manera colectiva, hay otro problema: bajo esa óptica, en la actualidad solo los muy poderosos (en especial, la farmaindustria) tendrían realmente la posibilidad de hacer aportes a la hora de valorar los fármacos. Lamentablemente, la justa demanda apuntada en la excelente nota de Teresa Forcades (6), según la cual los sistemas públicos de salud deberían desarrollar sus propios estudios independientes, no se vislumbra por ahora.

Es decir, la pertinencia de cierta magnitud del tamaño muestral se subordina casi siempre al dogma de la significación estadística.

Desde muy temprano se enfatizó (7) que cualquier diferencia observada diferirá de manera estadísticamente significativa de la nulidad siempre que el tamaño muestral sea suficientemente grande. Siendo así, sorprenden testimonios tales como que "es muy frecuente hallar artículos con tamaños de muestra insuficiente para detectar los efectos que valoran" (8), ya que, en rigor, esa no es que sea la situación "de muchos artículos" donde no se encuentra significación, sino la de absolutamente todos ellos.

Lo más atinado, por otra parte, sería emplear procedimientos que en lugar de convocarnos a desechar o no una hipótesis, los artículos se concentren en la estimación del efecto, tal y como sugieren desde 1988 las recomendaciones del *International Committee of Medical Journal Editors* (9) y con esa información "poner al día" la opinión que determinada hipótesis o punto de vista nos merece, a la luz de los nuevos datos.

En cualquier caso, lo más importante es que los resultados de un estudio pueden ser adecuadamente interpretados independientemente del tamaño muestral; lo verdaderamente necesario para ello es conocer las estimaciones realizadas –ya sea de los parámetros o de los efectos– con sus correspondientes intervalos de confianza.

REFERENCIAS BIBLIOGRÁFICAS

1. Ugalde A, Homedes N. Cuatro palabras sobre ensayos clínicos: ciencia/negocio, riesgo/beneficio. *Salud Colectiva*. 2011;7(2):135-148.
2. Silva LC. Diseño razonado de muestras y captación de datos para la investigación sanitaria. Madrid: Díaz de Santos; 2000.
3. Guyatt GH, Mills EJ, Elbourne D. In the era of systematic reviews, does the size of an individual trial still matter? *PLoS Medicine*. 2008;5:e4.
4. Malakoff D. Bayes offers a "new" way to make sense of numbers. *Science*. 1999;286:1460-1464.
5. Food and Drug Administration. Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials [Internet]. Silver Spring: FDA; 2010 [citado 10 de nov

2011]. Disponible en: <http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071072.htm>

6. Forcades i Vila T. Otras cuatro palabras clave para analizar críticamente un ensayo clínico: hipótesis, realización, análisis y publicación. [Debate]. *Salud Colectiva*. 2011;7(2):171-173.

7. Savage IR. Nonparametric statistics. *Journal of the American Statistical Association*. 1957;52:332-333.

8. Vickers AJ. Underpowering in randomized trials reporting a sample size calculation. *Journal of Clinical Epidemiology*. 2003;56(8):717-720.

9. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals [Internet]. 2010 [citado 6 oct 2011]. Disponible en: http://www.icmje.org/urm_main.html

FORMA DE CITAR

Silva Ayçaguer LC. Una pincelada estadística con repercusiones extrametodológicas. [Cartas]. *Salud Colectiva*. 2011;7(3):399-400.
