SALUD COLECTIVA, Buenos Aires, 7(3):399-400, September - December, 2011

A statistical brushstroke with extra-methodological repercussions

Una pincelada estadística con repercusiones extrametodológicas

Silva Ayçaguer, *Luis Carlos*[1]

[1]Researcher, Centro Nacional de Información de Ciencias Médicas. Tenured professor, Escuela Nacional de Salud Pública, La Habana, Cuba. Assistant Editor, Revista Cubana de Salud Pública. lcsilva@infomed.sld.cu

Dear Editor,

I would first like to highlight the importance the journal Salud Colectiva is gaining and to congratulate the collective that has made this growth possible. I have read with particular interest Volume 7 Issue 2, 2011, which in large part deals with the critical issue of clinical trials. One of its articles, entitled "Four words regarding clinical trials: science/profit, risks/benefit" (1), by the esteemed colleagues Antonio Ugalde and Núria Homedes, especially caught my attention. This article offers a significant contribution to the analysis of the modus operandi of the industry, ethics committees and regulatory bodies.

In the article, however, the authors manifest their agreement with a statement from a 2009 document of the US Food and Drug Administration (FDA), with which I dissent. The citation in question reads as follows:

> Another limitation is the sample size of the clinical trials. Because of economic issues and difficulties in recruiting patients who meet the inclusion criteria, a Phase III trial rarely includes more than 4,000 or 5,000 patients; vaccines trials are the exception, in which much larger samples made up of healthy individuals are used. A sample of 4,000 or 5,000 patients is not enough to represent the variety in genetics, sociodemographics (age, sex, sanitation conditions, etc.) and health conditions (concomitant diseases, nutritional state, etc.) of the population. (1 p.136)

Judgments regarding "insufficient sample size" continue to be firmly established evaluation criteria among ethics committees, research project evaluation agencies, and journal evaluators or editors. The legitimacy of these types of pronouncements is, however, highly debatable.

Such judgments are controversial and almost always inappropriate for at least two reasons. Firstly, because the formulas used to determine sample size are intrinsically speculative and entail an inevitable degree of subjectivity (2). Secondly, because the notion of a "sufficiently large size" only makes sense provided that operative rules or definite conclusions can be drawn from each isolated research work. However, this illusion is as deep-rooted as it is incorrect, for the simple reason that science just does not work that way.

Our scientific convictions may be more or less firm, but they are always provisional; while our representations of reality have at any given moment a certain degree of credibility, they are open to changes and improvements as suggested by new data. The consolidation of new knowledge is a gradual process in which any methodologically rigorous contribution is welcome. Some contributions will be more significant, others less so. Some will focus on certain age groups with certain social and public health conditions, and others on different groups and conditions. However, all studies can make a contribution to the ever-evolving process of consensus building, regardless of sample size. Progress is then made through meta-analysis (3) or the Bayesian approach (4), to mention the two best known alternative approaches.

The minimum sample size required by the ritualized cannons – obviously, the FDA is one of the organizations which follows these cannons, although it has recently taken a more nuanced position (5) – is one large enough to "detect" differences. This ambiguous expression presently refers

SALUD COLECTIVA, Buenos Aires, 7(3):399-400, September - December, 2011

to a size large enough to declare that the difference measuring the effect is statistically significant.

Apart from the fallacy implicit in the demand for huge trials, which ignores the collective nature of knowledge-building, there is yet another problem: if huge trials were truly necessary, only the most powerful (in particular, the pharmaceutical industry) would currently be able to make contributions to drug assessments. Unfortunately, the call made by Teresa Forcades in her excellent commentary to Ugalde and Homedes's article (5) – for public health systems to develop their own independent studies – is yet a distant possibility.

And so the relevance of a sample size of a given magnitude is almost always subordinated to the dogma of statistical significance. Early on it was stressed (7) that any observed difference will differ in a statistically significant way from the null as long as the sample size is sufficiently large. Therefore, it is surprising to find statements such as "one can frequently find articles with sample sizes insufficient to detect the effects they study"(8), because, strictly speaking, this is not the case only in the "many articles" where no significance is found, but in absolutely every one of them.

It would be wiser to use procedures which, instead of inviting us to consider whether or not to dismiss a hypothesis, focus on estimating effect; this has been suggested since 1988 in the recommendations of the International Committee of Medical Journal Editors (9). Articles structured in this way would thus help "update" the opinion we might have about a specific hypothesis or point of view in the light of the new data.

In any case, it is crucial to be able to adequately interpret the results of a research study regardless of the sample size; in order to do so it is necessary to know what estimates were made – either of the parameters or the effects – and their respective confidence intervals.

BIBLIOGRAPHIC REFERENCES

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

CITATION
Silva Ayçaguer LC. A statistical brushstroke with extra-methodological repercussions. [Letters]. Salud Colectiva. 2011;7(3):399-400.

The translation of this work is part of an interdepartmental collaboration between the Undergraduate Program in Sworn Translation Studies (English < > Spanish) and the Institute of Collective Health at the Universidad Nacional de Lanús. This article was translated by Adriana Coca and Lucas Traba, reviewed by María Victoria Illas and modified for publication by Vanessa Di Cecco.