Editorial

# On the interpretation of the evidence obtained by quantitative studies of health research

One of the aims of health research is to generate evidence on the effectiveness of treatments or preventive interventions. This type of problem, that requires comparing outcomes obtained or observed in two or more different groups of subjects, is usually addressed by quantitative research methods. The epistemological challenge is to determine whether the putative cause really generates the studied effect, i.e. whether the intervention caused the measured result or the studied factor is a cause of the observed outcome. This process of reasoning is called, in statistics and philosophy of science, *causal inference*.[1,2]

Causal inference is a process of inductive reasoning which leads to the conclusion, with varying degrees of certainty, that the estimates of the causal effects of a specific study are valid and generalizable.[3] For this to happen, it must meet a number of assumptions. Theoretically, a study is valid when neither random errors nor systematic errors (i.e. bias) exist.[4] Random errors

arise from the fact that quantitative studies are conducted using random samples of subjects. To minimize this error it should be ensured that the selection process is random and must have an appropriate sample size.[5] Biases are presented due to errors in the manner a research study is designed and operationally performed, which affect the comparability of data and results obtained from research subjects.[6] There may be biases in how the study subjects were selected (selection bias), bias in the way the data were collected (information bias, including measurement, memory and detection biases) and mixed biases among the former.[7,8]

Conceptually, one can estimate the causal effects of an intervention from experimental studies, because the control of the intervention by randomly assigning it can simulate a counterfactual scenario in which individuals under study are perfectly interchangeable in relation to their received treatment (intervention vs. control).[9] Thus, the *ideal experiment* is one that meets the following criteria: i) a large number of observation units (e.g. subjects) are randomly selected from a well-defined target population, ii) all known confounders are identified and measured, iii) allocation blocks are designed using those confounders, and iv) the treatment or the control is randomly assigned for each subject into those blocks.[2] But in real life, it is very difficult for a health research study to meet all the criteria of an ideal experiment.[4]

With regard to quantitative health research on individual subjects, it has been considered that the highest level of causal evidence is obtained by the *Randomized Controlled Trials* (*RCTs*, whether clinical, community or prevention trials) because theoretically the allocation of intervention is only determined by randomness and no other characteristics of the intervened subjects, or the context, explains the result; thus simulating a counterfactual experiment in which the outcome is explained only by the studied intervention.[6] With this type of experimental design, it should not exist treatment imbalance errors (i.e. differences in groups of subjects according to their assigned treatment) if the sample is large enough.[3] In addition, if the assessment of outcomes is performed masked regarding received treatment, the possibility of occurrence of information bias (i.e. systematic errors in measurement) is minimized.[10] Finally, to avoid bias in RCTs, two additional requirements are needed: that the study subjects comply with his/her assigned treatment according to the protocols and that follow-up losses do not occur.[6,11] Although RCTs are located at the top of hierarchy of evidence obtained by health research, just below the meta-analysis,[12] these trials have a number of disadvantages in relation to the ideal experiment, as defined previously: the selection of subjects participating in the trial is generally not performed at random, but for convenience (e.g. a group of patients suffering the disease in the hospitals of the study); moreover, this selection is also conditioned by inclusion and exclusion criteria of each study, and due to acceptance of participation by persons who were initially contacted. These selection problems lead to limitations in generalizing the results of RCTs beyond the participants' own characteristics, because the estimation error due to sample selection bias is not zero.[2,3] On the other hand, in most RCTs the sample sizes are very small and randomization is not performed in blocks, a situation that may generate residual estimation errors due to imbalances of subjects' characteristics (observed

and unobserved) among treatment groups, which could affect the study findings.[2,3] The latter problem can be partially solved with use of statistical adjustment methods for controlling confusion,[13] as multiple regressions;[4,14] but it should be kept in mind that the imbalance of treatments regarding unobserved variables is not resolved by these statistical analyzes;[4] thus under those circumstances, interpretation and generalization of RCTs findings will be limited.[3]

As explained, there are some warnings regarding interpretation of RCTs findings, but these cautions should be stricter when findings of analytical observational studies (i.e. cohorts, case-control, cross-sectional studies and their subtypes) are interpreted, because the process of causal inference from these studies is much more limited.[4,5] With the exception of studies based on random samples from well-defined populations, all observational epidemiological studies have some degree of estimation error due to the sample selection process (i.e. selection bias), which is much higher when samples are obtained for convenience (e.g. chosen from hospitalized control subjects) and no information is available to correct for the differential selection of study subjects (i.e. sampling weights or other characteristics of the selection process).[3] On the other hand, in most observational studies there is neither intervention nor treatment, but the effect of the studied exposure, or risk factor (e.g. use of cigarette, consumption of β-carotenes), which cannot be manipulated as it is in an experiment (i.e. RCTs in health research), instead it is observed as happened among the study subjects (e.g. smoking status).[7] Thus a phenomenon called confusion occurs; consisting on a spurious statistical association between an exposure variable (e.g. coffee consumption) and the outcome variable (e.g. lung cancer), which instead is really explained by a third variable, or a confounder (e.g. use of cigarette). The latter is related not causally with the exposure and is also independently associated with the outcome.[7,10] Consequently, there are estimation errors due to imbalances of characteristics (observed and unobserved) of exposed and unexposed subjects, affecting the study results.[3]

For the reasons explained here, in the analysis of observational studies it is not sufficient to estimate the bivariate relationships between exposures and outcomes (e.g. coffee consumption and lung cancer), instead more complex statistical methods are required, such as multiple regressions, matching or other methods not covered in the current editorial.[4,7,12]

Multiple regressions models allow for estimating the adjusted (or net) effect of exposure (or intervention) on the outcome variable, taking into account the effects of confounding variables.[4,13] To fit an appropriate regression model, prior knowledge about the studied phenomenon must be taken into account, in order to include the confounding variables indicated in each situation. These could be identified by techniques such as causal diagrams;[4,15] without including more variables than necessary in order to avoid an over-adjustment bias.[16] On the other hand, multiple regression models should avoid the collider-stratification bias, which occurs when a common effect variable (e.g. fever), which is caused by both the studied outcome (e.g. flu) and another study variable (or covariable, e.g. food poisoning), is included in the regression model rendering a spurious association between the covariable and the outcome (e.g. food poisoning and flu).[17] Despite these cautions, multiple regression models cannot control

for residual confounding due to unmeasured or unknown variables, which could bias the magnitude or the direction of study findings.[6,18]

On the other hand, matching is a method that allows a statistical balance between subjects who received the intervention (or exposure) and those who did not, by estimating the individual's propensity score of receiving the intervention, or not, given the characteristics of each individual before receiving that intervention (or exposure).[19] Using this score treated and untreated individuals are matched, seeking to balance the characteristics of people receiving and not receiving treatments, simulating the balance obtained when the interventions are randomly assigned, thus reducing the error due to nonrandom treatment allocation.[18,20] Finally, despite all the methods described here for designing and analyzing observational studies, when trying to estimate the effect of health interventions by means of observational studies, it is not possible to control adequately for *confounding by indication*, which refers to the inability to separate the effect of all indications of treatment (measured and unmeasured), from the actual treatment effect on the outcome variable.[18]

As described here, obtaining and interpreting results from quantitative studies, in order to generate health evidences, is a complex process that requires taking into account multiple and complex aspects related to the design and analysis of these studies, either experimental or observational. No research study in real life is perfect and no single study *per se* is sufficient evidence for changing health practices. Despite these limitations, when experimental and observational studies are conducted according to the current methods of health research,[4,7] both types of studies are complemented in the task of accumulating valid and reliable evidences, carried out by meta-analysis and systematic reviews,[21] in order to allow such evidences eventually being useful in improving the quality of life of the patients and their families.

**Andrés Fandiño-Losada MD, MSc, PhD.**
Assistant Professor and Coordinator,
Masters in Epidemiology.
School of Public Health.
Universidad del Valle, Cali, Colombia.

## References


1 Pearl J, Glymour M, Jewell NP. 2016 Causal Inference in Statistics: A Primer (Wiley, New York)

2 Rothman KJ, Greenland S. Causation and causal inference in epidemiology. Am J Public Health 2005; 95(S1): S144-S150.

3 Imai K, King G, Stuart EA. Misunderstandings between experimentalists and observationalists about causal inference. J R Stat Soc Series A (statistics in society) 2008; 171:481-502.

4 Rothman KJ, Greenland S, Lash T L(Eds.). 2008 Modern epidemiology.(Lippincott Williams & Wilkins)

5 Collins R, MacMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. The Lancet 2001; 357: 373-380.

6 MacMahon S, Collins R. Reliable assessment of the effects of treatment on mortality and major morbidity, II: observational studies. The Lancet 2001;357: 455-462.

7 Szklo M, Nieto J. 2012 Epidemiology. (Jones & Bartlett Publishers).

8 Delgado-Rodriguez M, Llorca J. Bias. J Epidemiol Community Health 2004; 58: 635-641.

9 Höfler M. Causal inference based on counterfactuals. BMC medical research methodology 2005 5(1): 1.

10 Gluud LL. Bias in clinical intervention research. Am J Epidemiol 2006;163:493-501.

11 Frangakis CE, Rubin DB . Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. Biometrika 1999;86:365-379.

12 Guyatt GH, Haynes RB, Jaeschke R Z, Cook DJ, Green L, Naylor CD, et al. Evidence-Based Medicine Working Group. Users' guides to the medical literature: XXV. Evidence-based medicine: principles for applying the users' guides to patient care. JAMA 2000;284:1290-1296.

13 de Irala J, Martínez-González M Á, Grima FG. ¿ Qué es una variable de confusión?. Medicina Clínica 2001;117:377-385.

14 McNamee R. Regression modelling and other methods to control confounding. J Occup Env Med 2005; 62:500-506.

15 Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. Am J Epidemiol 2002; 155:176-184.

16 Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. Epidemiology (Cambridge, Mass.) 2009; 20:488.

17 Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, Poole C. Illustrating bias due to conditioning on a collider. Int J Epidemiol 2010;39:417-420.

18 Bosco JL, Silliman RA, Thwin SS, Geiger AM, Buist DS, Prout MN, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. J Clin Epidemiol 2010;63:64-74.

19 Heinze G, Jüni P. An overview of the objectives of and the approaches to propensity score analyses. Eur Heart J 2011; 32:1704.

20 Stuart EA. Matching methods for causal inference: A review and a look forward. Statistical science: a review. Stat Sci 2010; 25: 1.

21 Shrier I, Boivin JF, Steele RJ, Platt RW, Furlan A, Kakuma R, et al . Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. Am J Epidemiol 2007; 166:1203-1209.