



Published in final edited form as:

*Obs Stud.* 2015 August ; 2015(7): 194–195.

## A good deal of humility: Cochran on observational studies

**Miguel A. Hernán**

Department of Epidemiology and Department of Biostatistics, Harvard T.H. Chan School of Public Health Boston, MA 02115, USA

Miguel A. Hernán: miguel\_hernan@post.harvard.edu

---

Cochran’s commentary shows that much of our conceptual framework for observational studies was already in place in the early 1970s. It also illuminates the technical progress that has been achieved since then, and identifies crucial methodologic challenges that still remain and that may remain forever.

In his commentary Cochran classifies observational studies into two groups. The first group—“analytical surveys” in Cochran’s terminology—investigates the “relation between variables of interest” in a sample of a target population. In these studies the goal is prediction, not causality. The second group studies the causal effects of “agents, procedures, or experiences.” Cochran’s commentary is almost exclusively concerned with this second group of observational studies whose goal is causal inference about comparative effects.

Cochran views these observational studies as attempts to answer causal questions in settings “in which we would like to do an experiment but cannot” because it is impractical, unethical, or untimely. The agents being compared in observational studies “are like those the statistician would call treatments in a controlled experiment.” Cochran effectively argues that observational studies for comparative effects can be viewed as attempts to emulate randomized experiments. Cochran, the Harvard statistician, was not alone. Other prominent researchers like Feinstein, the Yale epidemiologist, espoused similar views.

The concept of observational studies as an attempt to emulate randomized experiments was central for the next generation of statisticians and epidemiologists. Cochran argues that, like in randomized experiments, a prerequisite for causal inference from observational data is the statement of objectives or the “description of the quantities to be estimated.” Rubin, Cochran’s former student and future Chair of Statistics at Harvard, championed the use of counterfactual notation to unambiguously express these quantities as contrasts involving potential outcomes. A decade later Robins, also at Harvard, generalized counterfactual theory to time-varying treatments, a generalization that extends the concept of trial emulation to settings in which treatment strategies are sustained over time. These formalizations had profound effects on the field of causal inference from observational data.

A practical consequence of Cochran’s viewpoint is that observational studies can benefit from the basic principles that guide the design and analysis of randomized experiments. His commentary reminds us that causal analyses of observational data, like those of randomized trials, need to specify the “sample and target populations.” When discussing the “comparative structure”, he reminds us that studies with a control group, whether they are

observational or randomized, are generally preferred to those without a control group: “Single group studies are so weak logically that they should be avoided whenever possible”. And he identifies the defining problem of causal inference from observational data when two or more groups are compared: “How do we ensure that the groups are comparable?” This is the fundamental problem of confounding or, in Cochran’s terminology, “bias due to extraneous variables”.

Cochran classifies methods for confounding adjustment into three classes: analysis of covariance, matching, and standardization. In the decades after Cochran’s commentary, each of these methods was deeply transformed from simple techniques that could only be used under serious constraints (few covariates, linear models...) to powerful analytical tools with few restrictions. The analysis of covariance morphed into sophisticated outcome regression models that can easily handle complexities such as repeated measures, random effects, flexible dose-response functions, and failure time data. Matching was further developed for application to high-dimensional applications, which can incorporate propensity scores (co-developed by Rubin). Standardization in its two modern forms—the parametric g-formula and inverse probability weighting of marginal structural models (by Robins)—can now be applied to complex longitudinal data with multiple time points and covariates. In addition to these three classes of methods for confounding adjustment, a fourth class emerged in the early 1990s: g-estimation (Robins again). In many settings, the above methods can be made doubly-robust, another technical development that arose at the turn of the century. Finally, a whole suite of econometric methods like instrumental variable estimation are being progressively embraced by statisticians and epidemiologists interested in causal inference.

All these technical and conceptual developments, however, do not alter Cochran’s take home message: causal inference from observational data “demands a good deal of humility.” Fancy techniques for confounding adjustment will not protect us from bias if the confounders are unknown or if key variables of the analysis are mismeasured. Cochran reminds us that those who aspire to make causal inferences from observational data “must cultivate an ability to judge and weigh the relative importance of different factors whose effects cannot be measured at all accurately.” Because human judgment and subject-matter knowledge are fallible, causal inference from observational data is also fallible in ways that causal inference from ideal randomized experiments is not. A fascinating question is how much machine learning algorithms will be able to replace subject-matter knowledge in the years to come. For the time being, however, expert knowledge continues to be as paramount for the design and analysis of studies based on observational data as it was in Cochran’s time.