PROJECT MUSE®

Comment on Cochran's "Observational Studies"

Herbert L. Smith

Observational Studies, Volume 1, Number 1, 2015, pp. 217-219 (Article)

Published by University of Pennsylvania Press

DOI: <https://doi.org/10.1353/obs.2015.0023>

➔ *For additional information about this article*

<https://muse.jhu.edu/article/793422/summary>

Comment on Cochran's "Observational Studies"

Herbert L. Smith

hsmith@pop.upenn.edu

*Department of Sociology and Population Studies Center
University of Pennsylvania
Philadelphia, PA 19104, USA*

Cochran's perspective on observational studies (*inter alia*, Cochran 1965, Cochran 1972) remains a touchstone for population studies and other fields in which observations are plenty and experiments are implausible. It repays careful study and reflection. Cochran (1972, pp. 77-78) starts by distinguishing "analytical studies" from the "type of observational study [that] is narrower in scope[, where t]he investigator has in mind agents, procedures, or experiences that may produce certain *causal effects*...on people" [my emphasis]. The essay discusses difficulties that arise when treatments cannot be assigned at random, as in an experiment. It also discusses many other features of research design that have nothing to do with random assignment but much to do with our warrant for causal inference, including the need to be explicit regarding what the research is about (the specification of what is to be estimated and where effects are supposed to obtain), trade-offs in measurement between accuracy and bias, sample size, and non-response. Somewhere along the line "doing causation" became associated with establishing the internal validity of a study and/or form of estimation, and everything else got swept into external validity and measurement—important, to be sure; but somehow not *as* important as "causal analysis." To re-read Cochran (1972) is to realize this should not have happened (Smith 2009b, 2013).

In the discussion of sample size, Cochran (1972, p. 87) distinguishes between "estimating a single overall effect of the treatment" and the alternative situation in which "the variation in effect with an x is of major interest." In the latter circumstance, randomization does not buy one out of the need to specify accurately what is being estimated and for whom (Smith 1990). The definition of treatment effects at the unit level (Rubin 2005) has made clear how contingent all causal inference is on varying qualities of the population. Many of the x s are unobservable and are confounded with selection into treatment such that familiar statistical estimators must be re-conceptualized as to pertain to causal effects only in certain slices of the population (Angrist, Imbens, and Rubin 1996; Morgan and Winship 2007, ch. 7).

The renewed emphasis on the population heterogeneity of treatment effects (Xie 2013) has paid great dividends in the social sciences. It was always suspected that "returns to education" may be exaggerated given that those who obtain college educations may be precisely the type of people who would do well financially in the absence of a degree; notorious counterfactuals such as Bill Gates and Steve Jobs have only burnished the argument for adjusting downward maintained causal effects to adjust for positive selectivity into higher education. Although this may indeed be the case at the upper-end of the social background and ability distribution, there is a concomitant negative selection such that those most likely

to benefit from a college education are those least likely to be able to or to be encouraged to seek one (Hout 2012).

Cochran (1972, p. 87) wrote that “[i]n modern studies, standards with regard to the nonresponse problem seem to me to be lax.” The same is doubtless true in our post-modern era, but the problem has become so much worse that is hard to disentangle lax standards in survey research from social change (technologies, household and family structures, trust in institutions, community adhesion, non-stop marketing, and just plain fatigue with what was once an innovation but is now a hovering presence). Moreover, a blind faith in what are essentially arbitrary standards for response rates can distract from serious studies of bias due to non-response (Asch, Jedrzejewski, and Christakis 1997). We are learning interesting things. Early polling failures with respect to predicting presidential election outcomes have long been the basis for introducing students to problems of coverage and non-response bias alike (Freedman, Pisani, and Purves 1978, pp. 301-307). There are now so many polls, executed at varying degrees of departure—typically large—from the ideal of sampling frames isomorphic with target populations, respondent selection with known probabilities, and successful effort to deter non-response; and so many elections; that we know that non-response, though large in the extreme, is not a problem with respect to bias (Gelman and King 1993, pp. 423-427). Happily, all sorts of biases seem to cancel one another out, to the point where a mass of such polls, undifferentiated with respect to the classic criteria for such observational studies, provide a firm basis for some impressive forecasting of election results (Linzer 2013). Similar results have also been obtained via post-stratification adjustment of polls with no pretense to sampling (Wang et al. 2015).

But there are no guarantees that it will be ever thus—the 2015 British parliamentary election appears to be a contemporary example of results defying polls (if not modeling of poll data). Moreover, there are few if any similar domains of survey research in which the extent of observation is so great and the criterion for validity so decisive. Our hunches and received wisdom may not suffice. In a mail-out, mail-back survey of over 100,000 registered nurses, only one-third could be induced to respond. The possible extent of potential bias was large, and a careful follow-up survey of over 1,000 non-respondents (where only nine percent proved to be “hard core”) confirmed differential non-response with respect to a number of demographic characteristics, including gender, race, national origin, and education. These are among the characteristics often available in sampling frames, hence the basis of post-stratification weighting schemes. Yet the follow-up survey of non-respondents also revealed that in spite of the bias with respect to these demographic factors there was no bias whatsoever regarding the core content items that had motivated the study (Smith 2009a). I would be hard-pressed to dissent from Cochran’s (1972, p. 89) conclusion, that “observational studies are an interesting and challenging field which demands a great deal of humility, since we can claim only to be groping toward the truth.”

References

- Angrist, J.D., Imbens, G.W., and Rubin, D.B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91, 444-455.
- Asch, D.A., Jedrzejewski, M.K. and Christakis, N.A. (1997). Response rates to mail surveys published in medical journals. *Journal of Clinical Epidemiology*, 50, 1129-1136.

- Cochran, W. G. (1965). The planning of observational studies. *J. Roy. Statist. Soc. Ser. A*, 128 : 234-66.
- Cochran, W.G.. (1972). Observational studies. In *Statistical Papers in Honor of George W. Snedecor*, ed. T.A. Bancroft, 70-90. Iowa University Press, Ames. Reprinted in *Observational Studies*, 1, 126-136.
- Gelman, A. and King, G. (1993). Why are American presidential election campaign polls so variable when votes are so predictable? *British Journal of Political Science*, 23, 409-451.
- Hout, M. (2012). Social and economic returns to college education in the United States. *Annual Review of Sociology*, 38, 379-400.
- Linzer, D.A. (2013). Dynamic Bayesian forecasting of presidential elections in the states. *Journal of the American Statistical Association*, 108, 124-134.
- Rubin, D.B. (2005). Causal inference using potential outcomes. *Journal of the American Statistical Association*, 100, 322-331.
- Smith, H.L. (1990). Specification problems in experimental and nonexperimental social research. *Sociological Methodology*, 20, 59-91.
- Smith, H.L. (2009a). Double sample to minimize bias due to non-response in a mail survey. PSC Working Paper Series, December. http://repository.upenn.edu/psc_working_papers/20.
- Smith, H.L. (2009b). Causation and Its Discontents. In *Causal Analysis in Population Studies*, ed. H. Engelhardt, H.-P. Kohler, and A. F'urnkranz-Prskawetz, 233-42. The Springer Series on Demographic Methods and Population Analysis 23. Springer Netherlands. http://link.springer.com/chapter/10.1007/978-1-4020-9967-0_10.
- Smith, H.L. (2013). Research design: Toward a realistic role for causal analysis. In *Handbook of Causal Analysis for Social Research*, ed. S.L. Morgan. 45-73. Springer Netherlands. http://link.springer.com/chapter/10.1007/978-94-007-6094-3_4.
- Wang, W., Rothschild, D., Goel, S. and Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting*, 31, 980-991.
- Xie, Y. (2013). Population heterogeneity and causal inference. *Proceedings of the National Academy of Sciences*, 110, 6262-6268.