PROJECT MUSE®

Comment on Cochran's "Observational Studies"

Donald B. Rubin

Observational Studies, Volume 1, Number 1, 2015, pp. 212-216 (Article)

Published by University of Pennsylvania Press

DOI: <https://doi.org/10.1353/obs.2015.0022>

➔ *For additional information about this article*

<https://muse.jhu.edu/article/793421/summary>

Comment on Cochran's "Observational Studies"

Donald B. Rubin

rubin@stat.harvard.edu

Department of Statistics

Harvard University

Cambridge, MA 02138, USA

First, I have to thank Dylan Small for inviting me to contribute comments on the reprinted Cochran (1972) target article. This is, in fact, the third time that I have read this nice, almost colloquial, chapter Bill Cochran wrote in honor of his coauthor, George Snedecor. The first time was about 1970 when I was finishing my PhD under Bill's direction, and he circulated it as a pre-print. The second time was while I was writing my own chapter (Rubin 1984) summarizing Bill's contributions to observational studies, which appeared in the volume edited by Rao and Sedransk; because my discussion of it is less than three pages and the original is in a relatively massive book, I include it here in the appendix.

But what would I say that's different today than I said back in 1984? Obviously, I still am awed by Bill's straightforward and no-nonsense style of communication – that hasn't changed. But I think that today I would include some points not emphasized in 1984.

One aspect that I think I should have emphasized is the usefulness of the formal idea for an "assignment mechanism" to distinguish between randomized experiments versus observational studies and the formal concept of "potential outcomes" to define precisely causal effects. That is, under the "stable unit treatment value assumption" (SUTVA; Rubin 1980, 1986), $Y_i(1)$ are the i th units values of outcomes under the active treatment and $Y_i(0)$ are the i -th unit's values of outcomes under the control treatment, where the causal effect of the active versus the control treatment for unit i is the comparison of these two potential outcomes. Also, the assignment mechanism is the probability distribution of the vector of treatment indicators, $\mathbf{W}$, given the arrays of potential outcomes and covariates; this perspective was termed "Rubin's Causal Model" by Holland (1986), but the potential outcomes notation had its roots in the work of Neyman (1923) in the context of randomized experiments; the term "assignment mechanism" and the use of potential outcomes to define causal effects in general originates with work in the 1970s (Rubin 1974, 1976, 1977, 1978).

In my 1984 discussion of Cochran (1972), I did not emphasize the clarity that this conceptualization brings to causal inference in observational studies. In hindsight, I think that omission was because that conceptualization seemed so obvious to me. It is only in recent years that I have been befuddled by all the confusion created by some writers who eschew this formulation with its attendant clarity. The recent text by Imbens and Rubin (2015) hopefully contributes to rectifying this situation, at least from my perspective.

Another noteworthy omission in my 1984 discussion is my recent focus on the importance of outcome-free design for observational studies (Rubin 2006, 2008). I am not alone in having this current emphasis; see, for example Yue (2006) and D'Agostino and D'Agostino (2007). In hindsight I wished that I had emphasized this aspect, although with generous

interpretation one could read that theme into parts of Cochran (1972), although I do not see much distinction made between things like propensity score design, which is blind to outcome data, and model-based adjustment methods, which require outcome data and so are subject to inapposite manipulation. I think that this desire to correct this omission arises from being repeatedly exposed to more problematic examples in recent years.

A final comment on Cochran (1972) concerns a statement from his concluding section "Judgement About Causality" where he fairly blatantly reveals his disappointment in the answers provided by scientific philosophers. Often decisions about interventions must be made, even if based on limited empirical evidence, and we should help decision makers make sensible decisions under clearly stated assumptions so that "consumers" of the conclusion about the effects of some intervention can honestly weigh the support for that conclusion.

In conclusion: A fine choice for a classic article to reprint.

Appendix

Cochran (1972) is an easy to read article that begins with a review of examples of observational studies and notes their increased numbers in recent years. He mentions, as examples of observational studies for treatment effects, the Cornell study of seat belts (Kihlberg and Narragon, 1964), studies of smoking and health (U.S. Surgeon General Advisory's Committee Report, 1964), the halothane study (Bunker et al., 1969), and the Coleman report (Coleman et al., 1966). In many ways this article is an updated conversational summary of Cochran (1965). Here, even more than before, mature advice to the investigator is the focus of the paper.

First, the investigator should clearly state the objective and hypothesis of the study because such "...statements perform the valuable purpose of directing attention to the comparisons and measurements that will be needed." Second, the investigator should carefully consider the type of study. Cochran seems to be more obviously negative about studies without control groups than he was earlier:

Single group studies are so weak logically that they should be avoided whenever possible...

Single group studies emphasize a characteristic that is prominent in the analysis of nearly all observational studies – the role of judgment. No matter how well-constructed a mathematical model we have, we cannot expect to plan a statistical analysis that will provide an almost automatic verdict. The statistician who intends to operate in this field must cultivate an ability to judge and weigh the relative importance of different factors whose effects cannot be measured at all accurately.

Comparison groups bring a great increase in analytical insight. The influence of external causes on both groups will be similar in many types of study and will cancel or be minimized when we compare treatment with no treatment. But such studies raise a new problem – How do we ensure that the groups are comparable?

Cochran still emphasizes the importance of effective measurement:

The question of what is considered relevant is particularly important in program evaluation. A program may succeed in its main objectives but have undesirable side effects. The verdict on the program may differ depending on whether or not these side effects are counted in the evaluations...Since we may have to manage with very imperfect measurements, statisticians need more technical research on the effects of errors of measurement.

But the primary statistical advice Cochran has to offer is on controlling bias: “The reduction of bias should, I think, be regarded as the primary objective – a highly precise estimate of the wrong quantity is not much help...In observational studies, three methods are in common use in an attempt to remove bias due to extraneous variables...Blocking, usually known as matching in observational studies...Standardization (adjustment by subclassification)...Covariance (with x ’s quantitative) used just as in experiments.” The de-emphasis of efficiency relative to bias removal was evident when I began my thesis work under Cochran in 1968. The results of this thesis (Rubin, 1970) in large part published in Rubin (1973a, 1973b), and summarized in Cochran and Rubin (1973), lead to some new advice on the tradeoff between matching and covariance: In order to guard against nonlinearities in the regression of y on x , the combination of regression and matching appears superior to either method alone.

Recent work (Rubin, 1979) extends this conclusion to more than one x . Specifically, based on Monte Carlo results with 24 moderately nonlinear but parallel response surfaces and 12 bivariate normal distributions of x , and using percentage reduction in expected squared bias of treatment effect as the criterion, it appears quite clear that the combination of matched sampling and regression adjustment is superior to either matching or regression adjustment alone. Furthermore, Mahalanobis metric matching, which defines the distance between a treatment and control unit using the inverse of the sample covariance matrix of the matching variables and then sequentially finds the closest unmatched control unit for each experimental unit, was found superior to discriminant matching, which forms the best linear discriminant between the groups and sequentially finds the closest unmatched control unit for each experimental unit with respect to this discriminant. Moreover, regression adjustment that estimates the regression coefficient from the regression of the matched pair y differences on the matched pair x differences is superior to the standard covariance-adjusted estimator, which estimates the coefficients from the pooled within-group covariance matrix.

Cochran goes on to offer advice on sample sizes, handling nonresponse, the use of a pilot study, the desire for a critical colleague in the planning stages, and the relation between sample and target populations.

It is not surprising that this article concludes with a short section called “Judgment About Causality.” Cochran’s views are somewhat more bluntly presented here than in previous writing:

It is well known that evidence of a relationship between x and y is no proof that x causes y . The scientific philosophers to whom we might turn for expert guidance on this tricky issue are a disappointment. Almost unanimously and with evident delight they throw the idea of cause and effect overboard...A claim of proof of cause and effect must carry with it an explanation of the mechanism

by which the effect is produced. Except in cases where the mechanism is obvious and undisputed, this may require a completely different type of research from the observational study that is being summarized. Thus in most cases the study ends with an opinion or judgment about causality, not a claim of proof.

Cochran closes with the standard advice to make causal hypotheses complex:

Given a specific causal hypothesis that is under investigation, the investigator should think of as many consequences of the hypothesis as he can and in the study try to include response measurements that will verify whether these consequences follow.

References

- Bunker, J. P. et al., eds. (1969). *The national halothane study*. Washington, D.C.: USGPO.
- Cochran, W. G. (1965). The planning of observational studies. *J. Roy. Statist. Soc. Ser. A*, 128 : 234-66.
- Cochran, W.G. and Rubin, D.B. (1973). Controlling bias in observational studies: a review. *Sankhya A*, 35, 417-446.
- Coleman, J. S. (1966). *Equality of educational opportunity*. Washington, D.C.: USGPO.
- D'Agostino, R.B. Jr. and D'Agostino, R.B. Sr. (2007). Estimating treatment effects using observational data. *Journal of the American Medical Association*, 297, 314-316.
- Holland, P. W. (1986). Statistics and causal inference (with discussion). *Journal of the American Statistical Association*, 81, 945-970.
- Imbens, G.W. and Rubin, D.B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, New York.
- Neyman, J. (1923). On the Application of Probability Theory to Agricultural Experiments: Essay on Principles, Section 9. Translated in *Statistical Science*, 5, 465-480, 1990.
- Rubin, D.B. (1970). The Use of Matched Sampling and Regression Adjustment in Observational Studies, Ph.D. thesis, Department of Statistics, Harvard University.
- Rubin, D.B. (1973a). Matching to remove bias in observational studies. *Biometrics*, 29, 159-183.
- Rubin, D.B. (1973b). The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics*, 29, 184-203. *Biometrics*
- Rubin, D.B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, 688-701.
- Rubin, D.B. (1976). Inference and missing data (with discussion). *Biometrika*, 63, 581-592.
- Rubin, D.B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics*, 2, 1-26.
- Rubin, D.B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics*, 6, 34-68.
- Rubin, D.B. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, 74, 318-328.

- Rubin, D.B. (1980). Discussion of “Randomization analysis of experimental data in the Fisher randomization test by Basu.” *Journal of the American Statistical Association*, 75, 591–593.
- Rubin, D.B. (1984). William G. Cochran’s contribution to the design, analysis and evaluation of observational studies. In *W.G. Cochran’s Impact on Statistics*, eds., P.S.R.S. Rao and J. Sedransk. Wiley, New York.
- Rubin, D.B. (1986). Which ifs have causal answers? (comment on “Statistics and causal inference” by P.W. Holland). *Journal of the American Statistical Association*, 81, 961–962
- Rubin, D.B. (2008). For objective causal inference, design trumps analysis. *Annals of Applied Statistics*, 2, 808–840.
- Rubin, D.B. and Waterman, R.L. (2006). Estimating causal effects of marketing interventions using propensity score methodology. *Statistical Science*, 21, 206–222.
- United States Surgeon General’s Advisory Committee Report (1964). *Smoking and Health*. U.S. Department of Health, Education and Welfare, Washington D.C.