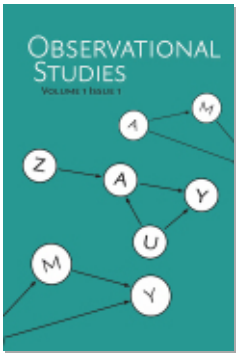
Comment on "Observational Studies" by Dr. W.G. Cochran (1972)

Mark J. van der Laan



Observational Studies, Volume 1, Number 1, 2015, pp. 220-222 (Article)

Published by University of Pennsylvania Press
DOI: *https://doi.org/10.1353/obs.2015.0024*

Observational Studies 1 (2015) 220-222 Submitted 3/15; Published 8/15

Comment on "Observational Studies" by Dr. W.G. Cochran (1972)

Mark J. van der Laan **laan@berkeley.edu**
Division of Biostatistics, School of Public Health
University of California, Berkeley
Berkeley, CA 94720, USA

This article "Observational Studies" by Dr. W.G. Cochran (1972) is not only a must-read but also an excellent article motivating the need for the *Observational Studies* journal that is now having its first issue.

Reading this paper by Dr. Cochran, a few points came to my mind. To start with his description of the growing importance of observational studies aiming to shed light on causal effects could have been written today with a few minor changes. For example, the importance of governments in driving the need for well designed observational studies, as he describes, is as timely as ever with the FDA running large scale safety analysis programs to determine harmful side effects of FDA-approved drugs, and with the precision medicine initiatives at the national and state level to improve patient care while making it cost-effective, to just name a few. These initiatives demand setting up large observational studies, such as the Sentinel Project, and create collaborations among government, insurance companies, industry and academics. By having a clear goal in mind, these initiatives provide unique opportunities for inspirational multidisciplinary bundles of scientific activities, all having the eye on the ball.

Dr. Cochran's articles represent the writing of a wise scientist that greatly cares about the real world, the truth, and the impact of these observational studies on society. The ending of his article is also telling about his great care: "In conclusion, observational studies are an interesting and challenging field which demands a good deal of humility, since we can claim only to be groping toward the truth."

Dr. Cochran provides a wonderful roadmap for planning observational studies, while providing crystal clear examples to demonstrate the dangers that lurk in the background and can completely destroy the success of an observational study. His demonstrations are as timely as ever, and the mistakes he warns against are not only still common place in the current era of Big Data, but are, in fact, even more prevalent than ever.

Even though much of his wisdom appears to be common place by now in a typical epidemiology or biostatistics education, he puts his finger on crucial spots that are easily overlooked by most practitioners and data analysts involved in designing observational studies, including its statistical analysis plan.

In particular, Dr. Cochran clarifies the importance of defining the causal quantity of interest and addressing what and how data needs to be measured in order to be able to learn the value of this causal quantity from the observed data. In addition, he stresses that once one has the question of interest in mind the decisions regarding the design can

now be targeted towards this question, resulting in what one might call targeted designs . An important part of our research has centered on precisely developing targeted group sequential designs that not only optimally estimate the desired causal quantity, but also adapt the design (e.g., allocation of treatment) towards its goal based on what one learned from past data (e.g., Chambaz et al, 2015), without sacrificing the scientific rigor of a controlled design.

Most importantly, time after time Dr. Cochran expresses concern about the underlying assumptions under which the design combined with the proposed analysis actually answer the question of interest. For example, he cares about the validity of a regression model, and that only with relatively few variables. I can only imagine how concerned he would be when we run a high dimensional linear regression analysis with hundreds or thousands of variables, as if these models approximate the truth. He emphasizes the need for supplementary studies, including pilot studies, to obtain better understanding of measurement errors "so that we can work with realistic models".

Clearly, Dr. Cochran does care about using statistical models that are realistic, and, even when the study is an observational study, he wants to control as much about the experiment as possible and incorporate this knowledge about the experiment in the statistical analysis plan for assessing the desired causal effect. Compare this with the common approach that throws one of our standard unrealistic parametric regression models at the data based on the format of the data, and manually tuning its choice to get the desired answer! In Dr. Cochrans discussion on the effect of misspecified models on bias and variance, he concludes with "Reduction of bias should be the primary objective". From this perspective, a recent opinion piece "Why we need a statistical revolution" (van der Laan, 2015) is very much in the spirit of Dr. Cochran. Our research in the field Targeted Learning (e.g., van der Laan and Rose, 2011) is aimed to respond to the enormous challenges our field is confronted with, by providing methods that optimally learn a specific target quantity, only incorporating real knowledge about the data generating experiment, fully utilizing the state of the art in machine learning through Super Learning, while still providing formal statistical inference. I would have loved the opportunity to talk with Dr. Cochran to hear his view on these robust methods based on realistic models, aiming to minimize bias, and maximize precision.

After having pointed out the lack of statistical methods that appropriately deal with nonresponse, Dr. Cochran concludes: "Fortunately, nonresponse can often be reduced materially by hard work during the study, but definite plans for this need to be made in advance." He realizes that observational studies require as much careful planning as a controlled experiment, and that hard work can prevent missingness or provide a fundamental understanding of the missingness mechanism so that statistical methods can correct for bias induced by informative censoring accordingly.

Beyond the great concern Dr. Cochran expresses regarding statistical bias due to violations of assumptions such as linearity of a regression model or measurement error assumptions, he pays particular attention to the non-testable assumptions required to draw causal conclusions. He states that these non-testable assumptions should not only be clearly presented and discussed in a separate section of a manuscript, but substantial effort should be invested in additional analyses that can shed some light on these assumptions, and possible explanations of the statistical findings should be provided.

Cochran strongly recommends the inclusion of a colleague who plays the role of devils's advocate to hit the weak spots of the statistical analysis plan. He would have liked the use of negative controls (i.e., in the actual data set of interest one assesses the effect of a variable on an outcome for which it is known that the causal effect equals zero) to showcase possible causal bias in the statistical method.

In light of the political maneuvering taking place in the current Big Data arena, the following remark by Dr. Cochran is very relevant and timely: "In numerous instances the choice seems to lie between doing a study much smaller and narrower in scope than desired but with high quality of measurement, or an extensive study with measurements of dubious quality. I am seldom sure what to advise."

To conclude this commentary, Dr. Cochran is one of the very important contributors to our discipline, and his spirit is at least as important now as it was at the time. His spirit stands for the advancement of science going after truth; careful and targeted planning of observational studies; targeting of the statistical approach towards the scientific question of interest and integrating the knowledge about the experiment; and hard work combined with humility when it comes to drawing conclusions. I am convinced that this new journal *Observational Studies* will stand for all this, greatly advance our scientific discipline, and thereby honor Dr. Cochran and the likes of him accordingly.

References

Chambaz, A. and van der Laan, M. J. (2014). Inference in targeted group-sequential covariate-adjusted randomized clinical trials. *Scandinavian Journal of Statistics*, 41, 104–140.

van der Laan, M.J. and Rose, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data.* Springer, New York.

van der Laan, M.J. (2015). Why we need a statistical revolution. `http://www.stats.org/super-learning-and-the-revolution-in-knowledge/`

Zheng, W., Chambaz, A. and van der Laan, M.J. (2015, forthcoming). Group sequential clinical trials with response-adaptive randomization. In *Modern Adaptive Randomized Clinical Trials: Statistical, Operational, and Regulatory Aspects*, ed. A. Sverdlov. Springer, New York. See also `http://biostats.bepress.com/ucbbiostat/paper323`