PROJECT MUSE®

Observational Studies and Study Designs: An Epidemiologic Perspective

Tyler J. Vander Weele

Observational Studies, Volume 1, Number 1, 2015, pp. 223-230 (Article)

Published by University of Pennsylvania Press

DOI: <https://doi.org/10.1353/obs.2015.0025>

➔ *For additional information about this article*

<https://muse.jhu.edu/article/793424/summary>

Observational Studies and Study Designs: An Epidemiologic Perspective

Tyler J. VanderWeele

tvanderw@hsph.harvard.edu

*Department of Epidemiology and Department of Biostatistics
Harvard University
Boston, MA 02115, USA*

Cochran's Contribution to Observational Studies

Cochran's remarkable paper, published over forty years ago, clearly articulates, or touches upon and anticipates, numerous principles of research for observational studies that are still of central relevance today. Amongst the principles and ideas covered in his paper are: the central importance of a control group (Section 3); issues of confounding and confounding adjustment (Sections 3 and 6); the relevance of baseline measurements of the outcome (Section 3); the issue of direction of causation even in what today would be called an interrupted time-series designs (Section 3); issues of measurement error including differential measurement error (Section 4); the importance of examining effects on multiple outcomes in assessing the relevance of a treatment or exposure (Section 4) - an issue that is still very much neglected today; randomized experiments as a template for thinking about observational studies (Section 5); matching, blocking, standardization, and covariance adjustment (Section 6); issues of generalizability and external validity (Section 7.5); and of course the challenges, and relevance, of causal inference (Section 7.6).

The discussion in Cochran's paper moreover anticipates numerous important ideas that were to be developed fully only much later. His discussion of matching and standardization and the difficulties encountered with multiple covariates (Section 6) anticipates the need for the use of propensity scores (Rosenbaum and Rubin, 1983, 1984) in the handling of these issues. His discussion of the role of blocking and covariance adjustment in observational studies to both increase precision and control bias, along with the position that in observational studies reduction of bias ought to be given priority (Section 5), partially anticipates Rosenbaum's notion of design sensitivity being more important than efficiency in observational studies (Rosenbaum, 2010). His discussion of causation as a change in the treatment variable leading to a change in the outcome (rather than merely being associated) anticipates the resurrection of Neyman's potential outcomes notation (Neyman, 1923) by Rubin (1974, 1978) for use in observational studies just a couple of years after Cochran's paper was written. His mention of "repeated before and repeated after" studies (Section 3), though not developed, anticipates in some sense Robins' development of concepts and methodology for the causal effects of time-varying exposures (Robins, 1986, 1997; Robins et al., 2000). Cochran's point that assessing the explanation of, or mechanisms for, an effect may require a completely different type of research from that assessing the overall effect (Section 7.6)

anticipates in some sense the literature on mediation and the challenges therein (Robins and Greenland, 1992; Pearl, 2001; VanderWeele, 2015). His encouraging the use of outcomes for which an exposure has no effect on the outcome (Section 7.6) anticipated more formal development of the use of “negative controls” for reasoning about causality (Lipsitch et al., 2010; Tchetgen Tchetgen, 2014; Imbens and Rubin, 2015). In many ways then Cochran’s paper provided a roadmap for a great deal of research on observational studies and causal inference that was to take place in the decades to follow.

Observational Studies and Study Design

Cochran’s discussion of observational studies is perhaps best understood as a statistician’s perspective. Many of the issues he had discussed are still at the forefront of statistical research on observational studies today. Within observational studies in epidemiology, however, while all of Cochran’s principles and discussion are still relevant, the scope of such studies is somewhat broader. Cochran distinguishes two broad types of observational studies: (i) analytic surveys and (ii) the studies of the effects of treatments. His focus, he notes, is on the latter. Implicit throughout nearly all of Cochran’s discussion of this second type of study is that the design of the study in question is what we might today call a cohort study. However, many other types of observational study designs are arguably also relevant in assessing the effects of treatments, and many of these designs have arisen within epidemiology. Indeed, as discussed below, one of epidemiology’s central methodological contributions has been to study design and to the development of new types of study designs. Several of these new designs are described below. However, before doing so, it will be good to consider more precisely what it is we mean by terms such as “study” and “study design.”

Cochran describes an observational study as a study in which “the investigator is restricted to taking selected observations or measurements on the process under study [and does not] interfere in the process in the way that one does in a controlled laboratory type of experiment.” Both (i) analytic surveys and (ii) studies of the effects of treatments fall within this description. His distinction above between these two might, however, be understood either as one of the data available (e.g. all measurements simultaneous versus some before and some after an exposure or event) or as one of purpose (descriptive versus causal). These two dimensions do not entirely coincide. In a survey in which all variables are measured on a single occasion, it may be possible to retrospectively assess certain exposures or treatments which can then be used in causal/etiologic research. Likewise, even with data available in a longitudinal cohort study, a simple description of the sample characteristics at baseline may, in some instances, be of considerable interest. We should thus distinguish between the data itself and how it was collected (study as a data resource) versus how the data is to be used (study as a particular empirical inquiry).

The term “study” might thus refer to either (i) “the process of systematically collecting data and the resulting data resource” or (ii) “the use of data to address a specific inquiry or set of related inquiries” (often resulting in a paper or report). Cochran’s use of “study” seems usually, though perhaps not always, to be in the latter sense. Even when he refers to specific studies by naming the data resource (e.g. the “National Halothane Study”), this is often in the context of a specific empirical inquiry. And indeed many studies (processes

of data collection and resulting data resource) are often focused upon addressing a single empirical question or set of closely related questions. This is especially the case with randomized trials in which the effectiveness of a specific intervention on a specific population is assessed as rigorously as possible. But, as will also be discussed further below, a single study (i.e. a single data resource) can often be used for multiple studies in the second sense i.e. for multiple empirical inquiries on different topics. When referring to a “study” it is thus good to distinguish between a “study” as the collecting of data resulting in a data resource and “study” as the steps taken to address a specific empirical inquiry or set of related inquiries.

Likewise “study design” might be understood either as (i) the design of the data collection process for a specific data resource, or (ii) the design of the analytic process and the use of that data resource to address a specific empirical question. Once again, most of Cochran’s discussion of the design of observational studies concerns study design in this latter sense. When the study (data resource) is put together for a focuses specific inquiry these two aspects of study design are then integrally intertwined. But, in other cases, the study (data resource) is put together for multiple purposes and can be used to address multiple very different inquiries; in these cases often these two aspects of study design may be quite distinct. Even after a study (data collection resulting in a data resource) has been conducted and the data is available, considerable work may be required in determining how that data resource can be used to address a specific empirical question.

The term “study design” does also have at least one yet further, and perhaps even more common use, that of a template. A specific “study design” might be understood as a template in which the actions undertaken to generate and collect the data fit a certain pattern or mold. A randomized trial might be viewed as one type of template, a closed cohort study as another type of template. Each of these is a type of study design. Thus, with the term “study design”, we have at least three uses: (i) the details of the design of the data collection process for a specific data resource (e.g. the design of the “National Halothane Study” or the “Framingham Heart Study”), (ii) the design of the analytic process used, including the data selected and the analyses conducted, to answer a specific empirical inquiry, and (iii) a template for obtaining data with a particular set of shared characteristics with respect to the data-generating mechanism. This third use of “study design” merits some further discussion, and in the following section several new types of study designs that have arisen in epidemiology are described.

Study Designs in Epidemiology

As indicated above, many different types of study designs (templates for the collection of data) can be used to assess causal/etiologic questions. Randomized trials and cohort type designs are perhaps the most common, but many other types of designs are available. In Section 4 of his paper, Cochran makes brief mention that cancer patients may be better informed of instances of cancer among blood relatives than are perhaps the controls who are free of cancer. Cochran is perhaps here touching upon the case-control design (though this is not entirely clear and a cohort design with case status ascertained at the end of follow-up, and perhaps certain covariate data ascertained retrospectively at that time, might also be in view). In any case, such case-control designs collect exposure/treatment data,

along with covariate data, on a set of individuals with the outcome, and also on a set of controls (selected either from the non-cases or from the underlying population). Intuitively, if the exposure has no effect on the outcome, the prevalence of the exposure should be no higher among the cases than among the controls. More specifically, in such case-control designs, if adequate control has been made for confounding, then it is possible to assess the effect of the exposure on the outcome at least on the odds ratio or rate ratio scale; one can obtain the same odds ratio or rate ratio from case-control data as one would have obtained in the underlying cohort (Miettinen, 1976; Prentice and Pyke, 1979). If further information is available on the prevalence of the outcome or the exposure, further progress can be made in estimating effects on the difference scale as well (Rothman et al., 2008). Such case-control studies are subject to additional biases, many of which have to do with selection of controls, but case-control designs can be very efficient ways to study cause-effect relationships especially when outcomes are rare. Such designs developed out of epidemiology, but there is no reason in principle why these designs could not be used in the social sciences as well, especially for rare outcomes for which larger sample sizes are needed when using cohort designs. Arguably their more extensive use in epidemiology is due to historic, rather than substantive, considerations. See Keogh and Cox (2014) for a recent comprehensive overview of case-control designs.

Epidemiologic study designs have also been extended in a number of other additional ways including varieties of case-control designs such as nested and matched case-control designs (cf. Rothman et al., 2008), two-stage sampling case-control designs (White, 1982; Breslow and Cain, 1988), and variants of the case-control design such as the case-cohort design (Prentice, 1986; Barlow, 1999), or two-stage designs involving both ecologic and case-control data (Haneuse and Wakefield, 2008). These other study designs could likewise have applicability outside of epidemiology. Even more creative, specialized, and in some way surprising designs have arisen within epidemiology that address questions of cause-effect relations. So called case-crossover designs collect data only on subjects who have the outcome in question and use exposure information in window of time substantially prior to the outcome (e.g. a week before the outcome) comparing this to exposure immediately preceding the outcome, to examine whether particular exposures triggered the outcome in question (Maclure, 1991; Maclure and Mittleman, 2000). Such designs have been used for example to study triggers for MI (Mittleman et al., 1993) and the relationship between cell-phone use and car accidents (Redelmeier and Tibshirani, 1997; McEvoy et al., 2005). They are useful for etiologic research but address a different type of causal question (e.g. does the exposure trigger the outcome in a short window of time following the exposure) than is addressed in cohort or case-control studies which focus on longer term effects (cf. Maclure, 2007). Such distinctions make clear the importance of specifying the object of investigation in etiologic research (Miettinen and Karp, 2012). It is not sufficient to simply specify the exposure and outcome being studied, but also the measure (prevalence, incidence; difference, ratio, rate, etc.) and the time-frame.

The case-crossover design described above is subject to biases that can arise from temporal trends in the exposure and various other types of designs such as the self-controlled case-series design (Farrington, 1995; Whitaker et al., 2006), the case-time-control design (Suijsa, 1995), and the case-case-time-control design (Wang et al., 2011) have been developed to help address some of these biases. This case-only type approach has also been

modified to study the effects of spatially relevant features or exposures, e.g. proximity to power lines that are only on one side of a street, by using data only on cases, and examining how distances to the feature would be modified if mirror images were taken with respect to some other spatial characteristic e.g. the center line of a street. Such designs are sometimes called case-specular study designs (Zaffanella et al., 1999). Yet another case-only study design variant arises from the observation, essentially simply a consequence of Bayes' Theorem, that if a genetic and an environmental exposure are independent in distribution in a population then the odds ratio relating the genetic and environmental exposures among the cases gives the measure of multiplicative interaction in the effects of the two exposures on the outcome in the population (Piergorsch et al., 1996; Schmidt and Schaid, 1999).

As with case-control designs, each of these other designs makes additional assumptions above and beyond those in a cohort study, and are thus subject to a wider range of biases (cf. Greenland, 1996; Rothman et al., 2008). But each of these designs also has the potential to give substantial insight in etiologic research, even though they are not traditional follow-up cohort studies and even though they sometimes do not even have a control group as traditionally conceived. In some instances, a study (data resource) produced for the purpose of one specific empirical inquiry can be useful in addressing an entirely different empirical study question by making use of one of the aforementioned, rather clever, study designs (templates). It is arguably the decoupling of the notion of study as data resource and that of study as template and as the use that is made of the data, that allows for such clever secondary uses of data, and that perhaps led to the development of these new study designs within epidemiology to begin with. Such decoupling has arguably also been important in trying to provide some unification of the different types of epidemiologic study designs (Miettinen and Karp, 2012). Once again, even though these designs were developed within epidemiology, there is generally nothing now restricting their use in other fields. Other disciplines would likely benefit from the use of these alternative observational study designs as well.

Large Multi-Purpose Cohort Studies

Even with cohort designs, the practice of epidemiology has arguably allowed for new insights. Cochran discusses the need for clear specific focused hypotheses and argues that a study should be designed in light of these (Section 2). It is hard to argue against such advice. However, if a study is conceived of as the creation of a data resource (e.g. the Nurses Health Study, the Framingham Heart Study etc.) then such a study can in fact be used to address multiple hypotheses. Hundreds of individual studies in the form of published empirical inquiries have come out of the Nurses Health Study on topics as diverse as nutrition, cancer, depression, social support, etc. In such a context, the principles of study design (understood as the design of the process by which the data resource is created and the resulting content of that resource) are arguably somewhat different than that for a study with a single narrow hypothesis. In the creation of a study (data collection resulting in a data resource) that is to be used for multiple research questions, careful thought must be given to the possibility of many hypotheses, what confounding variables are relevant not for just one exposure, or one outcome, but a whole host of exposure-outcome relationships; often extensive questionnaires and detailed follow-up to avoid non-response are needed to ensure adequate data.

The considerations when a single data resource is to be used for numerous empirical inquiries are challenging; but the undertaking of such studies (i.e. creation of such data resources) hold tremendous promise for research. Cochran, in his paper, discusses the difficult trade-off between conducting “a study much smaller...than desired but with high quality measurements, or an extensive study with measurements of dubious quality” (Section 4). In one of the rare moments of uncertainty in his paper, Cochran concludes: “I am seldom sure what to advise.” However, with large cohort studies that are to be used for multiple purposes, and which, often as a consequence, have extensive human and financial resources in support of them, there is sometimes no need to choose between the two. Large studies with excellent and extensive measurements are at least sometimes achievable. It is, once again, arguably the conceptual decoupling of “study” as the creation of a data resource and “study” as a specific empirical inquiry that underlies the pursuit of such large multi-purpose cohort studies.

Of course, even with a multi-purpose study (data resource), once the data has been collected, and a specific topic of study (empirical inquiry) has been selected, and an appropriate study design (template) chosen, a great deal of work may still remain in determining how best to make use of the data to answer the specific inquiry. And here, irrespective of the design (template) selected, all of Cochran’s principles come, once again, into play. Issues of selecting controls, baseline measurements, confounders and confounder adjustment, measurement error, method of adjustment, generalizability and external validity, and the assessing of the evidence for judging causation must all be carefully evaluated. Cochran’s discussion of observational studies is indeed as relevant today as it was forty years ago.

References

- Barlow, W.E., Ichikawa, L., Rosner, D., and Izumi, S. (1999). Analysis of case-cohort designs. *Journal of Clinical Epidemiology*, 52:1165-1172.
- Breslow N and Cain K (1988). Logistic regression for two-stage case-control data. *Biometrika*, 75:11–20.
- Farrington CP. (1995). Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics*, 51:228–235.
- Greenland S. (1996). Confounding and exposure trends in case-crossover and case-time-control designs. *Epidemiology*, 7(3):231-239.
- Haneuse, S.J.-P.A. and Wakefield, J.C. (2008). The combination of ecological and case-control data. *Journal of the Royal Statistical Society: Series B*, 70:73–93.
- Imbens, G.W., and Rubin, D.B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.
- Keogh, R.H., and Cox, D.R. (2014). *Case-control Studies*. Cambridge University Press.
- Lipsitch, M., Tchetgen Tchetgen, E. and Cohen, T. (2010). Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*, 21:383-388.
- McEvoy SP, Stevenson MR, McCartt AT, Woodward M, Haworth C, Palamara P and Cercarelli R. (2005). Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. *BMJ* 2005; 331.
- Maclure, M. (1991). The case-crossover design: a method for studying transient effects on the risk of acute events. *American Journal of Epidemiology*, 133:144-153.

- Maclure, M. (2007). “Why me?” versus “why now?” – differences between operational hypotheses in case-control versus case-crossover studies. *Pharmacoepidemiology and Drug Safety*, 16:850–853.
- Maclure M and Mittleman MA. (2000). Should we use a case-crossover design? *Annual Review of Public Health*, 21:193–221.
- Miettinen, O.S. (1976). Estimability and estimation in case-referent studies. *American Journal of Epidemiology*, 103:226–235.
- Miettinen, O.S. and Karp, I. (2012). *Epidemiological Research: An Introduction*. Springer.
- Mittleman MA, Maclure M, Toftler GH, Sherwood JB, Goldberg RJ and Muller JE. (1993). Triggering of acute myocardial infarction by heavy physical exertion. *New England Journal of Medicine*, 329:1677–1683.
- Neyman, J. (1923). Sur les applications de la thar des probabilités aux expériences Agricales: Essay des principe. Excerpts reprinted (1990) in English (D. Dabrowska and T. Speed, Trans.) in *Statistical Science*, 5:463–472.
- Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence*. Morgan Kaufmann, San Francisco, 411–420.
- Piegorsch, W.W., Weinberg, C.R., Taylor, J.A. (1994). Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Statistics in Medicine*, 13:153–162.
- Prentice, R.L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*, 73:1–11.
- Prentice, R.L. and Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika*, 66:403–411.
- Redelmeier DA and Tibshirani RJ. (1997). Association between cellular telephone calls and motor vehicle collisions. *New England Journal of Medicine*, 336: 453–458.
- Robins, J.M. (1986). A new approach to causal inference in mortality studies with sustained exposure period – application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7:1393–1512.
- Robins, J.M. (1997). Causal inference from complex longitudinal data. In: *Latent Variable Modeling and Applications to Causality*. Lecture Notes in Statistics (120), M. Berkane, Editor. Springer Verlag, New York, 69–117.
- Robins, J.M. and Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3:143–155.
- Robins J.M., Hernán M.A. and Brumback B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11:550–560.
- Rosenbaum, P.R. (2010). Design sensitivity and efficiency in observational studies. *Journal of the American Statistical Association*, 105:692–702.
- Rosenbaum P.R. and Rubin D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55.
- Rosenbaum, P.R. and Rubin D.B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79:516–524.
- Rothman, K.J., Greenland, S. and Lash, T.L. (2008). *Modern Epidemiology*, 3rd edition. Philadelphia: Lippincott Williams and Wilkins.

- Rubin, D. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, 66:688–701.
- Rubin, D.B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics*, 6:34–58.
- Schmidt, S., and Schaid, D.J. (1999). Potential misinterpretation of the case-only study to assess gene-environment interaction. *American Journal of Epidemiology*, 150:878–885.
- Suissa, S. (1995). The case-time-control design. *Epidemiology*, 6(3):248–253.
- Tchetgen Tchetgen, E.J. (2014). The control outcome calibration approach for causal inference with unobserved confounding. *Am. J. Epidemiol.* 179:633–640.
- VanderWeele, T.J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press, New York.
- Wang, S., Linkletter, C., Maclure, M., Dore, D., Mor, V., Buka, S. and Wellenius, G.A. (2011). Future-cases as present controls to adjust for exposure-trend bias in case-only studies. *Epidemiology*, 22(4): 568–574.
- Whitaker HJ, Farrington CP and Musonda P. (2006). *Tutorial in Biostatistics: The self-controlled case series method*. *Statistics in Medicine*, 25(10): 1768–1797.
- White, J. (1982). A two stage design for the study of the relationship between a rare exposure and a rare disease. *American Journal of Epidemiology*, 115:119–128.
- Zaffanella, L.E., Savitz, D.A., Greenland, S., and Ebi, K.L. (1998). The residential case-specular method to study wire codes, magnetic fields, and disease. *Epidemiology*, 9:16–20.