Potential pitfalls in observational study designs

Phillip Watkins, MS Statistics

Clinical research often uses observational data to compare two or more conditions to assess the viability of future experimental studies. A thorough literature review coupled with an understanding of the hierarchy of observational studies provides the first step needed to test a hypothesis (Figure). All observational studies related to a hypothesis should be completed before putting patients at risk in the planned experimental study. This review discusses the strengths and limitations of the various observational study designs and some common pitfalls to avoid.

A *case study* is a presentation of a single case that generates interest in the topic of study. The clinical values or an individual's characteristics are the only data to present. This is somewhat akin to a conversation one might have with a colleague, "I had the most interesting case the other day; have you ever seen anything like it?" The goal of publishing this information is to stimulate others to conduct studies with multiple similar cases.

A *case series* is a collection of case studies that probes for expected trends in future studies of this type of condition. As no control group is used in this study design, we usually compare descriptive statistics against standard values from healthy individuals taken from the medical literature. Resist the temptation to use a historical control, as the study group will differ with respect to time and likely location as well. Other problems can occur in reporting descriptive statistics for small sample sizes; if there are fewer than 30 continuous measurements, any means, standard deviations (SD) or 95% confidence intervals could be influenced by extreme values. To illustrate this, consider census data; the top 1% of earners skews the average income due to their inflated earnings. In such cases, a more robust statistic like the median and range (maximum - minimum) or a five number summary (minimum, Q1, median, Q3, maximum) along with boxplots can better summarize trends in these small, skewed data sets.

Sampling study with control data at a single point in time is commonly known as a cross-sectional design. *Cross-sectional* studies come in two varieties: descriptive and inferential. The *descriptive cross-sectional* study aims to estimate the *prevalence* of the condition in the two populations of interest or use the *correlation coefficient* to quantify the linear association between the suspected risk measure(s) and marker(s) of the disease. For example, one may compute the prevalence of childhood asthma in inner-city vs. suburban homes or compute the correlation between rescue inhaler uses in children vs. nearby carbon monoxide concentrations Remember that the relationship between the proposed risk factor and the disease level may not be linear, so this assumption should be confirmed with a scatter plot of the data.

In the *inferential cross-sectional* study, we test the hypothesis of differing prevalence of disease within the two groups or hope to show a statistically significant ($p < 0.05$) non-zero correlation between the risk factor and study outcome. Estimates obtained from the prior descriptive study will help power the study appropriately to ensure that the sample size gives a reasonable chance (typically 80%) at observing a statistically significant difference. Note that the cross-sectional design is relatively inefficient in comparing rare factors or outcomes, as a very large sample is needed before one expects to collect enough patients with the uncommon medical condition. Since most one time survey studies fall in this category, using a validated survey from the literature can help one

Corresponding author: Phillip Watkins
Contact Information: Phillip.watkins@ttuhsc.edu
DOI: 10.12746/swrccc 2016.0416.226

Figure 1: Algorithm for classication of types of clinical Research (courtesty of *The Lancet*).
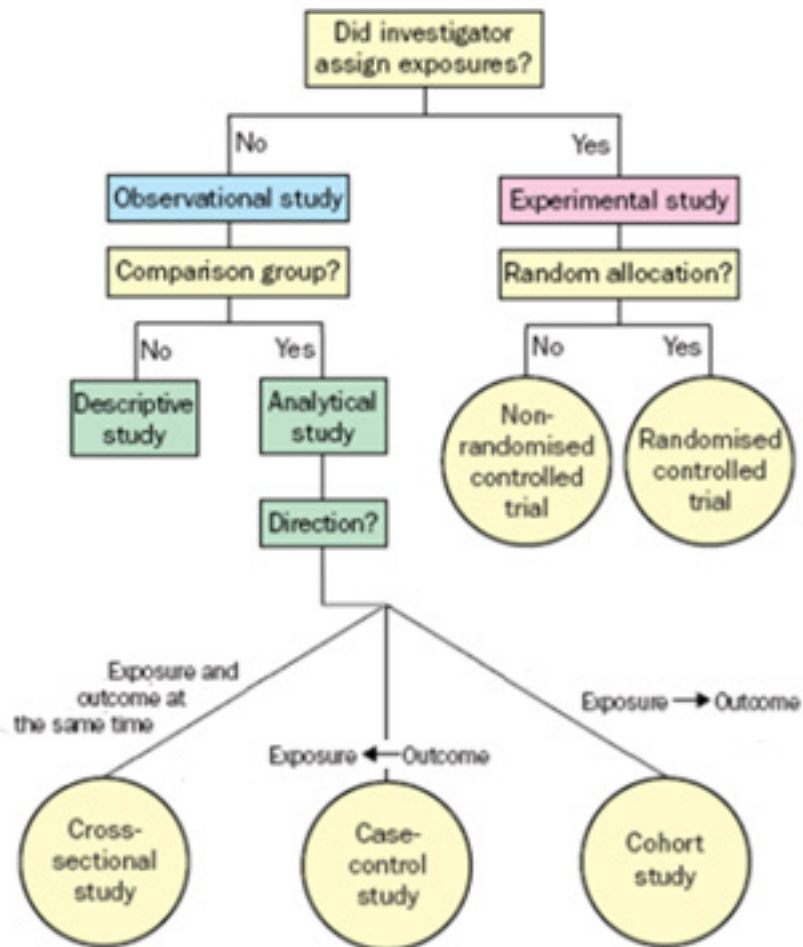
avoid a major pitfall, as a journal reviewer can reject a manuscript on the grounds that a "home-made" survey may not appropriately measure the condition(s) of interest.

The *case-control* study design compares cases with the disease from one population against controls from another population with respect to the risk factor(s) of interest. This design is much more useful for studying rare conditions, but is inefficient for studying rare exposures. By sampling the cases from one population and controls from another, there is always the potential for selection bias. This bias may cause secondary variables to confound the study results due to other incidental differences between our two groups. For example, patients with cancer may be found to be more likely to drink coffee until one adjusts for the association between coffee consumption and smoking. As such, it is crucial to compare baseline factors between the two groups and conduct the appropriate multivariate analysis to adjust for any observed clinical and/or statistically significant differences. Alternatively, a matched design may be used to make potentially confounding factors more comparable between the two groups and produce a more accurate odds ratio.

Use caution with odds ratios as they are often misinterpreted. For example, a case-control study of heart disease with a 1.5 odds ratio for smoking status implies that heart disease cases were 50% more likely to be smokers. To show that smokers were more likely to have heart disease, one must employ the subsequent cohort design. Also note that lack of specificity in "heart disease" and "smoker" status can invalidate either such study. Ideally, there should be clinically meaningful, pre-specified qualifiers for what constitutes "smoker" status, along with similar inclusion/exclusion criteria to adequately define "heart disease."

A *cohort study* tracks at risk and control individuals forward in time to compare the progression of the disease under study. Typically this study is conducted prospectively, though cohorts can be identified using retrospective data. While the latter method may save time, it does introduce additional bias and the investigator has less control over the nature of the study outcome measures. For example, people may not remember how many times they ate fast food last month, but they can certainly keep track for the next month! However, there is the danger of dropouts in prospective designs, so one should compare follow-up rates between the two groups at regular intervals during the study to detect this troublesome bias.

Just as case-control studies are inefficient for tracking rare exposures, cohort studies do a poor job evaluating rare outcomes. However, the major strength of tracking groups forward in time is to establish that the disease or condition occurs after the risk factor of interest. Showing that the risk precedes the disease is a necessary but not sufficient condition to show causation. In other words, while positive findings in a cohort study cannot establish causation, a sufficiently powered negative cohort study may contradict causation! For example, if stomach ulcers and dairy consumption are shown to be correlated in a case-control study, comparing ulcer rates in cohorts of milk and non-milk drinkers is likely to show that milk consumption does NOT increase the risk of ulcers.

Note that cohort studies also allow us to compute the incidence (new cases/study-years), relative risk, and various other useful comparative measures (risk difference, attributable risk %, etc.). As we usually think of time moving forward, the commonly reported relative risk is interpreted in a more intuitive fashion; a relative risk of 1.5 says that the at-risk group is 50% more likely to develop the disease or condition of interest. It is a common mistake to report an odds ratio (or adjusted odds ratio) under a cohort design instead of the relative risk, but this mistake is relatively harmless as the odds ratio approximates the relative risk when the sample is large or the exposure is rare.

In conclusion, the first step in conducting any study is a thorough review of the literature to determine what is already known. Determining which study design should come next in the project sequence; one can then design the appropriate study to reflect the desired analysis for publication. Consider using a checklist from STROBE (www.strobe-statement.org) as a template to ensure that no details are overlooked in planning your observational study. If you plan to test a hypothesis, seek out a statistician to help power your study design appropriately. Finally, an experimental study should come after all observational study designs have been conducted with definitive, positive results. Experimental studies have their own sets of pitfalls, which will be covered in a follow-up article.

Author Affiliation: Phillip Watkins is a statistician who works in the Clinical Research Institute at Texas Tech University Health Sciences Center in Lubbock, TX.

References

1.　　Greenberg RS, et al. Medical Epidemiology. 4th ed. New York, NY: Lange Medical Books/McGraw-Hill; 2001.

2.　　Grimes DA, Schulz KF. An Overview of clinical research: the lay of the land. The Lancet 2002; 359: 57-61. http://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(02)07283-5.pdf. Accessed July 1, 2016.

3.　　Röhrig B, et al. Study design in medical research. Dtsch Arztebl Int. 2009 Mar; 106(11): 184–189. http://dx.doi.org/10.3238%2Farztebl.2009.0184. Accessed July 1, 2016.

4.　　The STROBE initiative. Bern, Switzerland. http://www.strobe-statement.org/index.php?id=available-checklists. Accessed July 1, 2016.