Paper published in the *Journal of the Korean Data & Information Science Society*

"Simple power analysis in causal mediation models for a dichotomous outcome based on the mediation proportion"

Young Min Kim, John B. Cologne, Harry M. Cullings
Journal of the Korean Data & Information Science Society 2017 (May); 28(3):669–84
(doi: 10.7465/jkdi.2017.28.3.669)

Study Findings

Statistical methodology for studying complex causal models is developing rapidly, but one area requires further methodological development—the design of studies to detect mediation of radiation risk with a dichotomous (binary) outcome※. We used a computer-intensive approach to calculate power and sample size by using Monte Carlo simulation※※ and bootstrap methods※※※ for designing studies of disease prevalence with a continuous mediating variable. We also illustrated the utility of the method with simulation studies and an application to radiation risk of heart disease with serum HbA1c as a potential mediator (i.e. hypothetically related causally to radiation exposure).

※dichotomous (binary) outcome: an outcome that takes on only two possible values—e.g. yes or no (whether or not the event under study, such as a cancer, occurred in a study participant)

※※Monte Carlo: a term used to represent a random process in statistics; Monte Carlo simulation involves using a computer program to simulate random sets of data to aid in studying the statistical properties of a method, where "random" applies to natural variability in observed data, not the underlying parameters defining the data (such as the mean exposure in the population or the effects of radiation, which are fixed, not random)

※※※bootstrap methods: a set of statistical procedures that is based on the observed data values only, rather than being based additionally on assumptions about the underlying theoretical source of the data (such as its statistical distribution)

Explanation

Traditional epidemiological studies have relied on classical methods of regression analysis (linear regression※, logistic regression※※, or Cox regression※※※), which only test and estimate the association between a putative cause (explanatory variable: e.g. an exposure) and an effect (or outcome: e.g. cancer). In multivariate regression analyses with more than one explanatory variable, although the explanatory variables can be correlated, they cannot be causally related to one another, which is often not the case in practice. In other words, complex associations among study variables are often left untreated in classical regression analyses.

Recent calls to open this "black box" of inference have led to the expanded use of complex causal models in which there can be multiple regression relationships, including regression relationships among explanatory variables. Of particular interest is the situation in which an exposure, *X*, causes both the outcome, *Y*, and another explanatory variable, *Z*. In this case, part (or perhaps all) of the effect of *X* on *Y* might be explained by the pathway $X \rightarrow Z \rightarrow Y$, and *Z* is thus called a "mediator." In this case, interest often lies in the proportion of the total effect of *X* that is mediated by *Z*, or the "mediation proportion."

The foundation of such modeling lies in so-called "path analysis," which utilizes standard linear regression models for continuous explanatory variables and a continuous outcome variable. However, a straightforward path-analytic approach does not produce valid estimates of the mediation proportion when the outcome variable is not continuous, as happens when *Y* is a dichotomous (or binary: 1/0 or yes/no) variable representing the prevalence of disease, for example. In such studies, logistic regression is often used as the model for the effects of *X* and *Z* on *Y*; more theoretical work is needed to understand how to design studies to achieve adequate power for testing mediation and estimating the mediation proportion with such data.

※linear regression: the simplest form of statistical regression relationship, whereby an outcome variable (which is continuous, taking on many possible values, such as blood pressure) is assumed to follow a one-to-one straight-line (linear) relationship with a predictor variable (such as radiation dose)

※※logistic regression: a regression method that is suitable for dichotomous (binary) outcomes, for which there cannot be a one-to-one relationship between the predictor variable and the outcome variable, but a one-to-one relationship can be assigned between the predictor and the probability of the outcome occurring

※※※Cox regression: a regression method that is suitable for time-to-event outcomes (e.g. age at which a cancer occurs); although the ultimate outcome is dichotomous (e.g. whether or not a cancer occurred in a participant), the outcome variable used in the analysis is more informative, being not just the fact that an event occurred but also the amount of time it took for the event to occur

1. Study purpose

The purpose was to extend methodology to allow calculation of power and sample size in the case of a complex causal model involving mediation when the outcome is dichotomous.

2. Study methods

We applied computationally intensive Monte Carlo simulation and bootstrap methods to calculate power and sample size, and we evaluated the procedure by using simulation.

3. Study results

We successfully implemented the procedure and demonstrated its utility. We also showed that a study of heart disease prevalence with radiation effect mediated by serum HbA1c level could be conducted within the RERF Adult Health Study of atomic bomb survivors with greater than 80% power if the proportion of total radiation risk mediated by HbA1c were considered to be 0.33.

Study Significance

It is important to elucidate more precisely the mechanisms relating radiation exposure to disease. In addition to contributing to better scientific understanding of radiation effects, this could also result in clinically useful leads regarding long-term treatment following radiation exposure, if intervention were possible for the mediators. It is therefore crucial that methods be developed and implemented to allow assessing radiation risks with complex causal models incorporating mediation. The present work has extended our ability to perform such causal modeling in RERF studies of radiation risk.

The Radiation Effects Research Foundation has studied A-bomb survivors and their offspring in Hiroshima and Nagasaki for around 70 years. RERF's research achievements are considered the principal scientific basis for radiation risk assessment by the United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) and for recommendations regarding radiation protection standards by the International Commission on Radiological Protection (ICRP). RERF expresses its profound gratitude to the A-bomb survivors and survivors' offspring for their cooperation in our studies.

§***Journal of the Korean Data & Information Science Society***, which is the official journal of the Korean Data & Information Science Society, publishes original research articles, summary reports, reviews, case studies, and other invited papers in statistics, mathematics, and other related fields where relevant data analysis techniques are applied. (Impact factor in 2015: 0.808)