Editorial

Overview and practice of causal inference in observational studies

Editorial

 Causal inference has been well studied in Statisticsm[1] Epidemiology[2] and Social Science.[3] In order to conduct causal inference, three variables always need to be clearly defined: confounders, exposure variable and outcome variable. Confounders are usually patient information measured at the onset of a study, the exposure variable is a putative cause, and the outcome is an important endpoint of a research interest that determines the quality of a research program. Researchers have long been debating about ways to conduct causal analysis. In particular, the potential outcome framework[1] and causal DAG (Directed acyclic graphs)[4] have been most re-presentable conceptual modes of conducting causal inference in the 21st century. Nonetheless most clinical studies are still unfamiliar with the usage of such advanced causal inference methods and exclusively embrace of the randomization method, which is regarded as the gold standard of estimating the causal effect of an exposure condition such as a new intervention or novel drug. Thus this editorial gives an overview of the two causal inference methods with potential outcomes and causal DAG from the point of randomization.

a. Suppose for a moment that the randomization of certain exposure conditions was successful. Here a successful randomization procedure means that study subjects were fully compliant to assigned exposure conditions. Because the randomization procedure was successful, the exposure variable was not to be correlated with unmeasured confounders so that the effect of the exposure variable was to be unbiased. Now the connection between randomization and causal inference with the potential outcome framework is as follows: potential outcomes are the outcomes which one study subject can potentially observe under different exposure conditions. Consider an example of smoking and its effect on a lung cancer outcome. One can have his unique lung cancer outcome under smoking condition and another outcome under nonsmoking condition. The true causality is to compare these two potential lung cancer outcomes of the one same study subject. In reality, however, it is not possible to do so because only one of the potential outcomes is observed. That is, if one chooses to smoke, his/her lung cancer outcome under nonsmoking condition is not observed. This is called the fundamental problem of the causal inference.[5] How then does randomization solve this problem? Suppose for a moment that smoking could be successfully randomized despite its ethnical issue and that there were 50 smokers and 50 nonsmokers. Because of the successful randomization of the smoking condition, at least two valuable properties were accomplished: The two groups became comparable to each other with respect to confounders

b. The condition of being a smoker or nonsmoker became completely at random, equally likely to occur. Especially the second property means that the observed potential outcomes of the 50 smokers represent the entire 100 study subjects' potential outcomes that they would have if they had been smokers. Therefore, if randomization is successful, a simple comparison among the observed potential outcomes is the causal effect. In other words, randomization helps simple two-sample test report causal inference results. In causal DAG, successful randomization means that there are no arrows

Joseph Kang
Department of Preventive Medicine, Northwestern University, USA

Correspondence: Joseph Kang, Department of Preventive Medicine, Northwestern University, Feinberg School of Medicine, USA, Tel 7738502141, Fax 3129089588, Email joseph-kang@northwestern.edu

from confounders to the exposure variable. This means that regardless of conditions of confounders, the exposure variable influences its related outcome marginally and unbiased. Thus randomization ensures both the potential outcome framework and casual DAG to attain unbiased causal effect.[2]

 But in reality and many cases, the randomization of smoking conditions is unethical and impossible. Indeed there are many reasons why there are a vast number of observational studies especially in health sciences. If randomization is not successful, the effect is confounded, associated or correlated by confounders. Namely, the comparison of the observed outcomes among the groups defined by different exposure conditions is biased because the groups differ not only by the exposure conditions but also by confounders. Thus it is necessary to attain or approximate the ideal state of randomization. Over the past decades, statistical researchers have developed methods to shape an observational dataset to the dataset that could have been randomized. Such methods use propensity scores to conduct matching [6], stratification [7], and weighting [2]. The propensity score is defined as the probability of being exposed given measured confounders. The propensity score is a scalar summary of multidimensional measured confounders and has been used for matching, stratification and weighting with the unified goal of shaping observational data to be pseudo-randomized data. Matching is conceptually intuitive because subjects under different exposure conditions are matched as long as they share similar propensity scores. Stratification creates strata where subjects share similar propensity scores. Weighting enables observation data to have the exposure variable disassociated from measured confounding factors. Consider an example of a two-by-two table: smoking (yes or no) and obesity (yes or no). Regardless of cell frequency values of this two-by-two table, the odds ratio of this table will become 0.5 if they are weighted by the inverse of the propensity of smoking given the obesity condition.

 The weighted dataset represents and approximate a dataset that could have been randomized. More detailed description of such a weighting technique is described in Robins' seminal paper. In Epidemiology, such a weighting method is called the IPW (Inverse probability-weighted) method and is at the heart of the Marginal Structural Model.[2] All these methods have been well developed in R software packages: package Match it for matching and stratification and package twang for the IPW method.

There are a few caveats of practicing such causal inference methods. First, matching (except full matching) and weighting methods change the description of an observational dataset. Because those methods reshape the observational data, the descriptive statistics of reshaped data need to be reported along with that of the original data. Second, even though an observational dataset can be reshaped by these methods to attain balance among the measured confounders, this does not mean residual confounding is avoided. It is because only measured confounders are used to reshape the observational data, not unmeasured confounders. But there are methods to conduct sensitivity analysis for the unmeasured confounders.[8]

Conclusion

Causal inference with observational studies requires careful treatment of their datasets when matching, stratification or weighting with the propensity score are used. It is not impossible to draw causal conclusions in observational studies using those methods but the changed description of datasets by them must be reported.

Acknowledgements

None.

Conflicts of interest

Authors declare that there are no conflicts of interests.

References

1. Rubin D. Causal inference using potential outcomes: design, modeling, decisions. *Journal of the American Statistical Association*. 2005;100(469):322–331.

2. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550–560.

3. Morgan SL, Winship C. Counterfactuals and Causal Inference: Methods and Principles for Social Research, 1st edn. UK: Cambridge University Press; 2007. pp. 328.

4. Pearl J. Statistics and causality: separated to reunite–commentary on Bryan Dowd's "separated at birth". *Health Serv Res*. 2011;46(2):421–429.

5. Holland PW. Statistics and Causal Inference. *Journal of the American Statistical Association*. 1986;81(396):945–960.

6. Rosenbaum PR. Observational studies, 2nd edn. New York, USA: Springer; 2002. p. 377.

7. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*. 1984;79(387):516–524.

8. Brumback BA, Hernan MA, Haneuse SJ, et al. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Stat Med*. 2004;23(5):749–767.