Confounding in Observational Studies Evaluating the Safety and Effectiveness of Medical Treatments

Magdalene M. Assimon[1]

Introduction

Randomized controlled trials (RCTs) are considered the "gold standard" for establishing the safety and efficacy of medical treatments, such as drugs, devices, and procedures. Patients with kidney disease are often excluded from these studies (1), and it is well established that trial participants tend to be healthier than the broader kidney disease population (2). Furthermore, the number of nephrology-specific trials conducted continues to lag behind other subspecialties (3).

In the absence of RCT data, nephrology practitioners may look to population-specific observational evidence to guide therapy selection. Observational studies using real-world data (*e.g.*, administrative claims and electronic healthcare record data) to evaluate the safety and effectiveness of medical treatments can provide highly generalizable and valuable information to clinicians (4). However, like nonrandomized prospective cohort studies, these studies may suffer from biases that limit their validity, such as confounding.

In this commentary, I describe what confounding is and provide a brief overview of common types of confounding that can arise in observational studies of medical treatments. I then highlight some common strategies for addressing confounding and discuss potential sources of residual confounding.

Confounding

In an observational study, confounding occurs when a risk factor for the outcome also affects the exposure of interest, either directly or indirectly. The resultant bias can strengthen, weaken, or completely reverse the true exposure-outcome association. For a factor to be a confounder, it has to be associated with both the study exposure and the study outcome, and temporally precede the exposure (*i.e.*, it cannot be an intermediary factor on the causal pathway between the exposure and the outcome) (5).

Confounding by Indication and Examples of Other Types of Confounding

Confounding by indication (6) is one of the most common forms of bias present in observational studies evaluating the safety and effectiveness of medical treatments. It occurs when the clinical indication for treatment, such as the presence of a disease or disease severity, also affects the outcome of interest. Bias due to confounding by indication can make it appear that a treatment under investigation is associated with the occurrence of an outcome that it is supposed to prevent, especially in studies comparing the use of a medical treatment with nonuse. For example, confounding by indication would likely be present in an observational study assessing the association between aldosterone antagonist use versus nonuse and mortality in patients with heart failure. In such a study, heart failure severity is an important confounder. Clinicians are more likely to prescribe an aldosterone antagonist to patients with more severe heart failure, and more severe heart failure is also a risk factor for death. If heart failure severity is not adequately controlled for, it may appear that the use of an aldosterone antagonist increases the risk of death, which is contrary to existing evidence from placebo-controlled trials (7).

Confounding by frailty (8) can be another important source of bias in observational studies of medical treatments. This type of confounding occurs because frail patients, who are close to death, tend to have a lower likelihood of receiving preventative therapies than individuals who are healthier. When confounding by frailty is present, the preventative treatment being evaluated appears to be more beneficial than it actually is. For instance, confounding by frailty has been proposed as a potential explanation for the implausible 40%–60% mortality reduction seen in observational studies assessing influenza vaccine effectiveness in older adults (9). Compared to healthier patients, frailer patients with a poor short-term prognosis may be less likely to receive an influenza vaccine due to a perceived lack of benefit. In this scenario, frailty is a confounder because it associates with vaccine receipt and death.

Other types of confounding can arise when heathy behaviors are associated with both the medical exposure under study and the outcome of interest. For example, confounding by the healthy adherer effect (10) occurs because patients who adhere to treatments tend to have a higher likelihood of taking part in other beneficial healthy behaviors (*e.g.*, exercising) than their nonadherent counterparts. When confounding by the healthy adherer effect is present, studies evaluating

University of North Carolina Kidney Center, Division of Nephrology and Hypertension, Department of Medicine, University of North Carolina School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

Correspondence: Dr. Magdalene M. Assimon, University of North Carolina Kidney Center, 7024 Burnett-Womack CB #7155, Chapel Hill, NC 27599-7155. Email: massimon@med.unc.edu

the effect of treatment adherence versus nonadherence on the occurrence of adverse clinical outcomes will often overestimate the beneficial effects of treatment adherence.

Finally, time-varying confounding occurs when the exposure of interest and potential confounders change across time. A common type of time-varying confounding that may be present in observational studies of medical treatments is "time-varying confounding affected by previous exposure." (11) It arises when the clinical parameter indicating that a treatment change is necessary is independently related to the outcome of interest and is also affected by previous exposure to the treatment (12). For example, in a study assessing the association between erythropoietin-stimulating agent (ESA) dose and mortality in patients on hemodialysis, serum hemoglobin is a time-varying confounder that needs to be accounted for. Hemoglobin levels predict ESA dose, are influenced by prior ESA dose, and are independently associated with mortality (the outcome).

Strategies for Addressing Confounding

Confounding can be addressed in the design and analytic phases of observational studies. Common strategies are discussed below, and their advantages and disadvantages are summarized in Table 1.

Addressing Confounding in the Design Phase

Restriction is a method than can be used for confounding control in the design phase. Similar to RCTs, restriction in an observational study involves setting criteria for study inclusion. By limiting the study to individuals who meet specific criteria, confounding by each respective inclusion criterion is either eliminated or reduced. For instance, in an observational study evaluating the risk of fracture associated with the use versus nonuse of benzodiazepines, age and sex are likely important confounders. Restricting the study cohort to males who are <65 years of age would eliminate confounding by sex and reduce confounding by age. Confounding by sex is eliminated because there is no variation in benzodiazepine use by sex—all benzodiazepine users and nonusers are male. Limiting the study cohort to individuals <65 years of age does not completely remove confounding by age, because benzodiazepine use patterns and fracture risk likely varies across the 18- to 64-year-old age group. Although restriction is an intuitive method that can be easily implemented, potential disadvantages include sample size reduction and decreased generalizability.

Another confounding control strategy that can be used in the design phase is matching. In a cohort study, matching involves selecting a comparator group that is matched to the treatment group on one or more confounders. Usually, individual-level matching is performed. Consider the previously mentioned observational study evaluating the benzodiazepine-fracture association. Because age and sex are important confounders, one or more benzodiazepine nonusers would be matched to a patient taking a benzodiazepine on the basis of age and sex. For example, a 63-year-old female not taking a benzodiazepine would be matched to a 63-year-old female taking a benzodiazepine. Although exact matching on the basis of age is ideal, it may not be possible. Broader age-based matching categories—such as matching

on age within 5 years—can be used, but residual confounding by age may remain. In addition, it is important to keep in mind that identifying matched pairs of treated and comparator patients becomes more difficult as the number of matching factors increases.

Specific to observational studies evaluating medical treatments, a design strategy that can be used to minimize the effect of confounding by indication is using an active comparator rather than a nonuser comparator. The treatment of interest and the selected comparator should have the same clinical indication and therapeutic role, and in the case of medications, have the same mode of delivery (4). Furthermore, using an active comparator is the only logical comparator choice when irretractable confounding by indication is expected. Besides mitigating confounding by indication, head-to-head comparisons of two or more treatments with the same indication provide relevant information on comparative safety and effectiveness that can be used to inform the selection of one treatment over another in clinical practice.

Addressing Confounding in the Analytic Phase

There are several statistical approaches that can be used for confounding control in the analysis phase. Multivariable adjustment, which involves including potential confounders as covariates in regression models, is the most common analytic technique used. However, recently, propensity score methods, such as propensity score matching and propensity score weighting, have gained popularity (13).

In studies evaluating medical treatments, a propensity score is a patient's predicted probability of receiving the treatment of interest versus a comparator, given their measured baseline characteristics. This summary score is estimated for each patient in the study cohort and is subsequently used for confounding control. In propensity score matching, each patient who received the treatment of interest is matched to one or more patients who received the comparator with an equivalent propensity score. This results in the generation of a matched cohort of treated and comparator patients that have similar baseline characteristics. In propensity score weighing, the propensity score is used to generate weights that are applied to the original study cohort to create a pseudo-population of treated and comparator patients that have similar baseline characteristics (14). The resultant matched and weighted cohorts can be used to estimate the treatment-outcome association, where the influence of measured baseline confounding is minimized. Propensity score methods and multivariable adjustment typically yield similar adjusted estimates of the treatment-outcome association (13). However, because a propensity score combines multiple covariates into a single summary score, these methods are preferred when the exposure of interest is common and outcome of interest is rare, a setting where multivariable outcome models are susceptible to overfitting. Readers interested in learning more about propensity score methods can refer to the tutorial provided by Fu *et al.* (15).

G methods, such as inverse probability–weighted marginal structural models, are complex analytic methods that appropriately handle time-varying confounding in the setting of time-varying exposures. A thorough description of G methods is beyond the scope of this commentary and

Table 1. Advantages and disadvantages of common strategies used to address confounding

Method	Overview	Advantages	Disadvantages
Design phase			
Restriction	Setting criteria for study inclusion	Easy to implement	Only removes or reduces confounding by the inclusion criteria Reduces sample size Cannot generalize findings to those excluded
Matching	Creates matched sets of patients who have similar values of one or more confounders	Intuitive	Difficult to match on multiple confounders Only removes or reduces confounding by the matching factors Unmatched patients are excluded, reducing sample size, effect estimate precision, and generalizability
Active comparator	Comparing the treatment of interest to an active comparator rather than treatment nonuse	Mitigates confounding by indication Clinically relevant head-to-head comparison of two or more treatments	Cannot be used when there is only one treatment option
Analysis phase			
Multivariable adjustment	Potential confounders are included as covariates in regression models	Easy to implement in standard statistical software packages	Only controls for measured confounders The total number of confounders that can be included in regression models is contingent on the number of outcome events
Propensity score matching	Each patient who received the treatment of interest is matched to one or more patients who received the comparator treatment with an equivalent propensity score, generating a matched cohort of treated and comparator patients that have similar baseline characteristics	Preferred in studies where there are relatively few outcome events compared with the number of potential confounders Ability to check if covariate balance between the treated and comparator groups was achieved in the matched cohort	Only controls for measured confounders Unmatched patients are excluded, reducing sample size, effect estimate precision, and generalizability
Propensity score weighting	The propensity score is used to generate weights that are applied to the original study cohort to create a pseudo-population of treated and comparator patients that have similar baseline characteristics	Preferred in studies where there are relatively few outcome events compared with the number of potential confounders Ability to check if covariate balance between the treated and comparator groups was achieved in the weighed cohort	Only controls for measured confounders Less intuitive than propensity score matching
G methods	Complex analytic methods that handle time-varying confounding in the setting of time-varying exposures	Appropriately handle time-varying confounding	Only controls for measured confounders Complex methods requiring advanced statistical expertise

can be found elsewhere (11). However, it is important to recognize that the use of these methods is increasing in the field of nephrology.

Common Sources of Residual Confounding

Despite the use of study designs and analytic strategies that aim to eliminate confounding, residual confounding may persist. Common reasons why residual confounding may be present are: (1) information on a confounder is not available; (2) the version of the confounding variable present in the data source is an imperfect surrogate or is misclassified; and (3) continuous confounders are parameterized as categoric variables, especially when overly broad categories are used (16).

Conclusion

Observational studies using real-world data can provide clinically actionable information on the potential benefits and harms of medical treatments in populations excluded from RCTs, such as patients with kidney disease. Confounding is a common source of bias threatening the validity of these studies. Thus, it is important to be aware of the types confounding that may be present and understand the advantages and disadvantages of common strategies used for confounding control.

Author Contributions

M.M. Assimon wrote the original draft and reviewed and edited the manuscript.

References

1. Konstantinidis I, Nadkarni GN, Yacoub R, Saha A, Simoes P, Parikh CR, Coca SG: Representation of patients with kidney disease in trials of cardiovascular interventions: An updated systematic review. *JAMA Intern Med* 176: 121–124, 2016 10.1001/jamainternmed.2015.6102
2. Smyth B, Haber A, Trongtrakul K, Hawley C, Perkovic V, Woodward M, Jardine M: Representativeness of randomized clinical trial cohorts in end-stage kidney disease: A meta-analysis. *JAMA Intern Med* 179: 1316–1324, 2019 https://doi.org/10.1001/jamainternmed.2019.1501
3. Strippoli GF, Craig JC, Schena FP: The number, quality, and coverage of randomized controlled trials in nephrology. *J Am Soc Nephrol* 15: 411–419, 2004 https://doi.org/10.1097/01.ASN.0000100125.21491.46
4. Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM, editors: *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide*, Rockville, MD, Agency for Healthcare Research and Quality (US), 2013
5. Rothman KJ, Greenland S, Lash TL: *Modern epidemiology*, 3rd Ed., Philadelphia, Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008
6. Walker AM: Confounding by indication. *Epidemiology* 7: 335–336, 1996
7. Pitt B, Zannad F, Remme WJ, Cody R, Castaigne A, Perez A, Palensky J, Wittes J; Randomized Aldactone Evaluation Study Investigators: The effect of spironolactone on morbidity and mortality in patients with severe heart failure. *N Engl J Med* 341: 709–717, 1999 https://doi.org/10.1056/NEJM199909023411001
8. Glynn RJ, Knight EL, Levin R, Avorn J: Paradoxical relations of drug treatment with mortality in older persons. *Epidemiology* 12: 682–689, 2001 https://doi.org/10.1097/00001648-200111000-00017
9. Zhang HT, McGrath LJ, Ellis AR, Wyss R, Lund JL, Stürmer T: Restriction of pharmacoepidemiologic cohorts to initiators of medications in unrelated preventive drug classes to reduce confounding by frailty in older adults. *Am J Epidemiol* 188: 1371–1382, 2019 https://doi.org/10.1093/aje/kwz083
10. Shrank WH, Patrick AR, Brookhart MA: Healthy user and related biases in observational studies of preventive interventions: A primer for physicians. *J Gen Intern Med* 26: 546–550, 2011 https://doi.org/10.1007/s11606-010-1609-1
11. Mansournia MA, Etminan M, Danaei G, Kaufman JS, Collins G: Handling time varying confounding in observational research. *BMJ* 359: j4587, 2017 https://doi.org/10.1136/bmj.j4587
12. Bradbury BD, Gilbertson DT, Brookhart MA, Kilpatrick RD: Confounding and control of confounding in nonexperimental studies of medications in patients with CKD. *Adv Chronic Kidney Dis* 19: 19–26, 2012 https://doi.org/10.1053/j.ackd.2012.01.001
13. Stürmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S: A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol* 59: 437–447, 2006 https://doi.org/10.1016/j.jclinepi.2005.07.004
14. Brookhart MA, Wyss R, Layton JB, Stürmer T: Propensity score methods for confounding control in nonexperimental research. *Circ Cardiovasc Qual Outcomes* 6: 604–611, 2013 https://doi.org/10.1161/CIRCOUTCOMES.113.000359
15. Fu EL, Groenwold RHH, Zoccali C, Jager KJ, van Diepen M, Dekker FW: Merits and caveats of propensity scores to adjust for confounding. *Nephrol Dial Transplant* 34: 1629–1635, 2019 https://doi.org/10.1093/ndt/gfy283
16. Kyriacou DN, Lewis RJ: Confounding by indication in clinical research. *JAMA* 316: 1818–1819, 2016 https://doi.org/10.1001/jama.2016.16435