Confounding

K.J. van Stralen^a F.W. Dekker^b C. Zoccali^c K.J. Jager^a

^aERA-EDTA Registry, Department of Medical Informatics, Academic Medical Center, University of Amsterdam, Amsterdam, and ^bDepartment of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; ^cCNR-IBIM Clinical Epidemiology and Pathophysiology of Renal Disease and Hypertension, Renal and Transplantation Unit, Ospedali Riuniti, Reggio Calabria, Italy

Key Words

Epidemiology · Confounding · Nephrology · Bias ·
Confounding by indication · Causality · Associations

Abstract

In confounding, the effect of the exposure of interest is mixed with the effect of another variable. It is important to identify relevant confounders and remove the confounding effect as much as possible. There are three criteria that need to be fulfilled to determine whether a variable could be considered a potential confounder. The first criterion is that the variable needs to be associated with the exposure. The second criterion is that the variable needs to be associated with the outcome or disease. The third criterion is that the variable should not be an intermediate variable in the causal pathway between exposure and outcome. Only if all the criteria are fulfilled is the variable under question a confounder. If one incorrectly adjusts for a variable that is not a confounder, one risks overadjustment or adjustment for spurious associations. Confounders can be prevented from entering the study, during the design of a study, or if this is not possible, one can try to remove it during the analysis phase.

Copyright © 2010 S. Karger AG, Basel

The previous paper in this series focused on different types of bias, which was defined as a systematic error in the design or conduct of the study. Confounding is something different, although some consider it a special type of bias. The term confounding comes from the Latin *confundere*, to mix together, and refers to the ‘confusion’ or mixing of effects, i.e. the effect of the exposure of interest is mixed with the effect of another variable. This then leads to erroneous results. We will firstly describe how confounders can be identified and then deal with how to remove confounding in experimental and observational studies.

Assessing the Presence of a Confounder

Confounding has to do with trying to ascertain causality in observing exposure-outcome relationships [1]. If one would like to study whether hypertension is causally related to the development of end-stage renal disease (ESRD) [2], one expects that obesity will most likely affect this relationship, and the question arises if obesity is a confounder in this relationship between hypertension and ESRD. But how can we determine if obesity is a confounder?

There are three criteria that need to be fulfilled to determine whether a variable could be considered a poten-

Fig. 1. Schematic representation of obesity as a confounder in the relation between hypertension and ESRD.

Fig. 2. Schematic representation of obesity as a potential confounder in a study on lead exposure and the development of ESRD. As obesity is not related to lead exposure, obesity is not a confounder, but simply another risk factor for ESRD.

tial confounder. The first criterion is that the variable needs to be associated with the exposure. The second criterion is that the variable needs to be associated with the outcome or disease. The third criterion is that the variable should not be an intermediate variable in the causal pathway between exposure and outcome. Only if all the criteria are fulfilled is the variable under question a confounder. We will now illustrate this reasoning with some examples. In doing so we will show that one specific variable, in this case obesity, may be a confounder in some relationships but not in others.

In the Relationship between Hypertension and ESRD, Is Obesity a Confounder?

To determine if this is the case, we check the three criteria. First we know that obesity is associated with hypertension as many obese patients develop hypertension (criterion 1 fulfilled). Second we know that obesity is a factor that is related to ESRD, i.e. many obese patients develop ESRD (criterion 2 fulfilled) [2]. Finally, as hypertension does not cause obesity, it is not in the causal pathway (criterion 3 fulfilled). So, we could say that all the criteria are fulfilled and that obesity can be considered a confounder (fig. 1).

In the Relation between Lead Exposure and ESRD, Is Obesity a Confounder?

The situation, however, can be different for different relationships. In the relation between lead exposure in drinking water and the development of ESRD [3], we may also want to know if we need to consider obesity as a confounding factor. We know that obesity is related to the

development of ESRD (criterion 2 fulfilled). However, being obese is not related to lead exposure (criterion 1 fails). If a potential confounder is not related to the exposure, but only to the outcome, the potential confounder is simply a different independent risk factor for the disease, in this case ESRD. Therefore it will not be necessary to adjust for it (fig. 2).

In the Relation between Obesity and ESRD, Is Caloric Intake a Confounder?

When we study the relation between obesity and ESRD, we could wonder whether we would need to adjust for caloric intake. Of course, obesity has a strong relation with caloric intake, as a high caloric intake is a cause of obesity (criterion 1 fulfilled). However, a 'direct' relation between caloric intake and ESRD is much less likely (criterion 2 fails). We can reasonably assume that nearly all of the effect of caloric intake on the development of ESRD will work 'via' the pathway of obesity. Therefore if a potential confounder is related to the exposure, but has no 'direct' relation to the disease, the variable is simply a risk factor for or cause of the exposure, and it is not necessary to adjust for it (fig. 3).

In the Relation between Race and ESRD, Is Obesity a Confounder?

When we would perform a study on the relation between race as a risk factor for ESRD [4] we could wonder whether obesity is a confounder. Obesity rates vary between races (criterion 1 fulfilled), and obesity is related to the development of ESRD (criterion 2 fulfilled). However, obesity could be considered a consequence of race, result-

Fig. 3. Schematic representation of caloric intake as a potential confounder in a study on obesity and the development of ESRD. Obesity and caloric intake are related, but as caloric intake does not affect the risk of ESRD apart from obesity, caloric intake is not a confounder in the relation between obesity and ESRD.

Fig. 4. Schematic representation of obesity as a potential confounder in a study on race, and the development of ESRD. Race is associated with obesity, obesity is associated with ESRD. However, different races have differences in obesity rate. Therefore, they will have a different risk of ESRD via an increased risk of obesity, and obesity is in the causal pathway and is therefore not a confounder.

ing in the development of ESRD. Therefore, obesity is not a confounder, but an intermediate variable (criterion 3 fails). If a variable is in the ‘causal pathway’ between the exposure and the disease, this means that part of the relation between the exposure and the disease goes via this variable.

In the first few examples, if one would falsely assume that one would need to adjust for obesity in the relation between lead exposure and ESRD, or for caloric intake, this would not have large consequences on the effect estimate, although it may result in a loss of statistical power and the risk of adjusting for spurious associations. However, this is not true for variables that are in the causal pathway. If we mistakenly adjust for a variable in the causal pathway, we remove part of the effect of the exposure, namely all the effect that goes via the intermediate variable. One therefore overadjusts the effect, resulting in an incorrect reduced estimate of the effect of the exposure. So, in this example, one would falsely reduce the differences between races. In general, when studying the effect of race or (other) genetic association studies, confounding is not an issue, as most variables are either not related to race (or the gene studied), or they are in a causal pathway between race or the gene and the outcome.

Sometimes, one wants to make use of this ‘overadjustment’, although this is not for reasons of confounding but to explore (causal) mechanisms. For example, if one would like to know the relation between race and ESRD, *independent* of differences in obesity or if one would like

to study how much of the differences between races could be explained by obesity [5], one on purpose can ‘overadjust’ for obesity (fig. 4).

To decide whether a variable is a confounder is a matter of judgment, and therefore a decision by the investigator. Although variables can have a strong statistically significant correlation, it is important to discriminate between true confounding, spurious associations or variables that are intermediary in causal pathways [1]. Such decision-making cannot be determined via a statistical model, but needs profound knowledge on pathophysiology. As shown in the examples, obesity can be either another risk factor for the disease, an intermediate or a confounder, depending on the exposure-outcome relationship studied. Therefore, for each separate analysis one should determine what the potential confounders are.

One Step Further

Reasoning on potential confounding can be quite complicated. In the previous examples it was relatively straightforward to conclude whether a variable was a confounder or not. However, this is not always the case. An example: if we were to study exercise and the development of ESRD, we could discuss whether obesity is a potential confounder [6]. Assessing the three criteria for confounding, we can conclude that obesity is related to

Fig. 5. Schematic representation of obesity as a potential confounder in a study on exercise and the development of ESRD. However, the relation between obesity and exercise is complicated. On the one hand, obesity is partially a consequence of exercise and therefore it is in the causal pathway. On the other hand, the 'urge' to exercise could be considered a consequence of obesity, so therefore being also a confounder.

exercise (criterion 1 fulfilled) and the obesity is related to ESRD (criterion 2 fulfilled). However, the relationship between exercise and obesity is complicated. Some individuals become obese because they exercise too little. In this situation, obesity is in the causal pathway between exercise and ESRD (criterion 3 fails). Therefore, one could consider obesity not to be a confounder. Conversely, in others, being obese may make people more motivated to start exercising, or make individuals feel reluctant to exercise (criterion 3 fulfilled). In this latter situation, obesity is not in the causal pathway, and can be considered a confounder.

So what is true? Both reasonings are true. On the one hand, if we would adjust for obesity we would remove some of the effect of exercise as we overadjust for the intermediate variable. On the other hand, if we would not adjust, we miss adjusting for a confounding factor. Logic reasoning is crucial to determine what, in a specific study, would be more important and likely. That conclusion would be up to the researcher (fig. 5).

How to Deal with a Confounder?

As a confounder obscures the real effect of an exposure, one needs to remove its effect as much as possible. However, the effect of a confounder is present in the population and will therefore be found in all types of observational studies. Fortunately, there are multiple ways that

one could use to deal with a confounder. One can prevent confounding when designing a study; if this is not possible one can try to remove it during the analysis phase.

There are three methods used to prevent confounding in the design of epidemiologic studies. The first solution is to perform an experiment or randomized controlled trial [7]. In a randomized controlled trial the confounding factors are randomly distributed, by chance, over the exposure groups. Randomization is especially important when studying the effectiveness of a treatment. A physician prescribes often a treatment based on certain characteristics of a patient. Therefore, it is usually impossible to compare different treatments, as treatment groups are intrinsically different on many confounders and other variables. This is called 'confounding by indication'. The only solution to this problem is randomization of the different treatments. However, it is not always possible or ethical to perform a randomized controlled trial. Like in a study on obesity, it is impossible to have people being exposed or unexposed to being obese.

A second solution to confounding is restriction, i.e. to select only those individuals with the same value for the confounding factor, like selecting only non-obese individuals. It is a very efficient technique, but makes the results less generalizable to the general population. Furthermore, if selection criteria are very strict, like when selecting only subjects with a body mass index of 23, recruitment of eligible participants can be difficult.

The third solution in preventing confounding when designing a study is to match individuals on the presence of the confounding factor. In short, matching may be performed subject by subject, which is known as individual matching, or for groups of subjects, which is known as frequency matching [8]. One matches a (group of) exposed subject(s) with the confounding factor to another (group of) unexposed subjects(s) with the confounder, for example, taking the same number of obese and non-obese individuals in the exposed and non-exposed group. This latter method will be explained in a future paper. Like restriction, matching may make the results of a study less generalizable to the entire population.

Adapting your study design to a specific confounder is a very drastic measure, usually only suitable for a small number of confounders. However, studies often suffer from multiple confounders. Therefore, an easy solution is trying to remove the effect of the confounder during the statistical analyses. This can be done by stratification or multivariate analysis. These methods will be discussed in future papers.

Conclusion

Confounding is a significant threat to the evaluation of causal relationships. It is important to identify relevant confounders and remove the confounding effect as much as possible. One needs to define exposure-outcome rela-

tionships in advance and check the three criteria described to determine whether something is a confounder, otherwise one risks overadjustment or adjustment for spurious associations. The subject of confounding should therefore receive careful attention during the study design and the data analysis.

References

- 1 Vandenbroucke JP: The history of confounding; Morabia A (ed): A History of Epidemiologic Methods and Concepts. Basel, Birkhäuser, 2004.
- 2 Hsu CY, Iribarren C, McCulloch CE, et al: Risk factors for end-stage renal disease: 25-year follow-up. Arch Intern Med 2009;169:342–350.
- 3 Yu CC, Lin JL, Lin-Tan DT: Environmental exposure to lead and progression of chronic renal diseases: a four-year prospective longitudinal study. J Am Soc Nephrol 2004;15:1016–1022.
- 4 Tomson CR, Foley RN, Li Q, et al: Race and end-stage renal disease in the United States Medicare population: the disparity persists. Nephrology (Carlton) 2008;13:651–656.
- 5 Krop JS, Coresh J, Chambless LE, et al: A community-based study of explanatory factors for the excess risk for early renal function decline in blacks vs. whites with diabetes: The Atherosclerosis Risk in Communities Study. Arch Intern Med 1999;159:1777–1783.
- 6 Stengel B, Tarver-Carr ME, Powe NR, et al: Lifestyle factors, obesity and the risk of chronic kidney disease. Epidemiology 2003;14:479–487.
- 7 Stel VS, Zoccali C, Dekker FW, Jager KJ: The randomized controlled trial. Nephron Clin Pract 2009;113:c337–c342.
- 8 Rothman KJ, Greenland S: Modern Epidemiology, ed 2. Philadelphia, Lippincott-Raven, 1998, pp 1–737.