http://www.kidney-international.org

Confounding: What it is and how to deal with it

KJ Jager[1], C Zoccali[2], A MacLeod[3] and FW Dekker[1,4]

[1]ERA–EDTA Registry, Department of Medical Informatics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; [2]CNR–IBIM Clinical Epidemiology and Pathophysiology of Renal Diseases and Hypertension, Renal and Transplantation Unit, Ospedali Riuniti, Reggio Cal, Italy; [3]Department of Medicine and Therapeutics, University of Aberdeen Medical School, Aberdeen, UK and [4]Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands

As confounding obscures the 'real' effect of an exposure on outcome, investigators performing etiological studies do their utmost best to prevent or control confounding. Unfortunately, in this process, errors are frequently made. This paper explains that to be a potential confounder, a variable needs to satisfy all three of the following criteria: (1) it must have an association with the disease, that is, it should be a risk factor for the disease; (2) it must be associated with the exposure, that is, it must be unequally distributed between exposure groups; and (3) it must not be an effect of the exposure; this also means that it may not be part of the causal pathway. In addition, a number of different techniques are described that may be applied to prevent or control for confounding: randomization, restriction, matching, and stratification. Finally, a number of examples outline commonly made errors, most of which result from 'overadjustment' for variables that do not satisfy the criteria for potential confounders. Such an example of an error frequently occurring in the literature is the incorrect adjustment for blood pressure while studying the relationship between body mass index and the development of end-stage renal disease. Such errors will introduce new bias instead of preventing it.

Kidney International (2008) **73,** 256–260; doi:10.1038/sj.ki.5002650; published online 31 October 2007

KEYWORDS: confounding; randomization; matching; stratification; epidemiology

Correspondence: *KJ Jager, ERA–EDTA Registry, Department of Medical Informatics, Academic Medical Center, J1b 115.1, PO Box 22700, Amsterdam 1100 DE, The Netherlands. E-mail: k.j.jager@amc.uva.nl*

Received 12 June 2007; revised 27 July 2007; accepted 7 August 2007; published online 31 October 2007

Confounding, sometimes referred to as confounding bias, is mostly described as a 'mixing' or 'blurring' of effects.[1] It occurs when an investigator tries to determine the effect of an exposure on the occurrence of a disease (or other outcome), but then actually measures the effect of another factor, a confounding variable. As most medical studies attempt to investigate disease etiology and causal relationships, confounding is regarded as undesirable, as it obscures the 'real' effect of an exposure. For this reason, confounding is something that investigators want to get rid of, for example, by so-called 'adjustment for confounding variables'. This paper will explain the concept of confounding as well as the ways in which confounding can be addressed, including randomization, restriction, matching, and stratification. Another common way to address confounding, multivariate analysis, will be discussed in future articles in this series.

WHEN ARE VARIABLES POTENTIAL CONFOUNDERS?

To explain the phenomenon of confounding, it is necessary to consider the relationship between an exposure and the occurrence of a disease (Figure 1). In order for a variable to be a potential confounder, it needs to have the following three properties: (1) the variable must have an association with the disease, that is, it should be a risk factor for the disease; (2) it must be associated with the exposure, that is, it must be unequally distributed between the exposed and nonexposed groups; and (3) it must not be an effect of the exposure, nor (linked to this) be a factor in the causal pathway of the disease.

This can be illustrated by a study on the relationship between dialysis modality at the start of renal replacement therapy (RRT) and patient survival (Figure 2) as was performed by Couchoud *et al.*[2] in Example 1.

Example 1. Association between treatment choice and outcome in the elderly with end-stage renal disease (ESRD) *Couchoud et al.[2] studied the association between initial dialysis modality and 2-year patient survival in a cohort of 3512 elderly ESRD patients. After adjustment for estimated glomerular filtration rate (estimated GFR) at dialysis initiation and a number of other factors, unplanned HD was associated with a 50% increased risk of death and PD with a 30% increased risk of death compared with planned HD.*

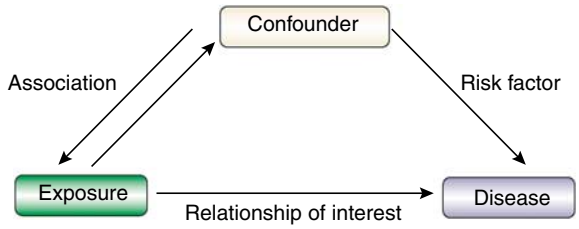

Figure 1 | Properties of a confounder.

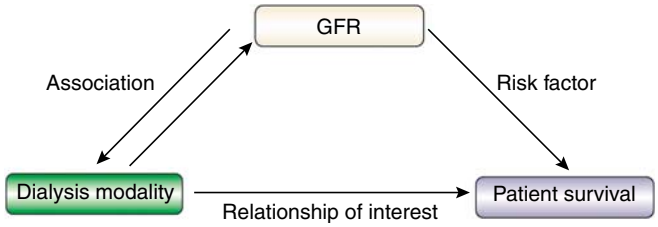

Figure 2 | Relationship between dialysis modality and patient survival—could GFR immediately prior to the start of dialysis be a confounder in this relationship?

In this study, Couchoud *et al.*[2] rightfully adjusted for estimated GFR because it is known from the literature that the GFR prior to the start of dialysis is a risk factor for patient survival on dialysis. Second, it is known that, on average, patients starting on peritoneal dialysis tend to have a higher GFR than those starting on hemodialysis; hence, there is an association between dialysis modality and GFR. Finally, GFR prior to the start of dialysis may influence the choice of dialysis modality, but is certainly not an effect of dialysis modality, as it was measured before initiating RRT. Thus, the third requirement for GFR to act as a confounder has also been fulfilled. It can therefore be concluded that GFR prior to the start of dialysis is indeed a potential confounder in the relationship between dialysis modality and patient survival.

A second example relates to the relationship between body mass index (BMI) and the development of ESRD (Figure 3). This relationship was studied by Hsu *et al.*[3] in Example 2.

Example 2. BMI and the risk for ESRD *Hsu et al.*[3] *investigated the relationship between BMI and the risk for ESRD using data of more than 320 000 members of Kaiser Permanente. They were able to show that, adjusted for a number of confounders like age, sex, and race (but not for blood pressure), increased BMI was strongly associated with an increased risk for ESRD.*

Reviewing the different properties of potential confounders, it can be stated that blood pressure is indeed a risk factor for the development of ESRD. Furthermore, there is an association between BMI and blood pressure, in that higher BMIs are associated with higher blood pressures. Blood pressure is, however, also an effect of BMI; it is probably even in the causal pathway that leads to the development of ESRD. Hsu *et al.*,[3] therefore, rightfully concluded that blood

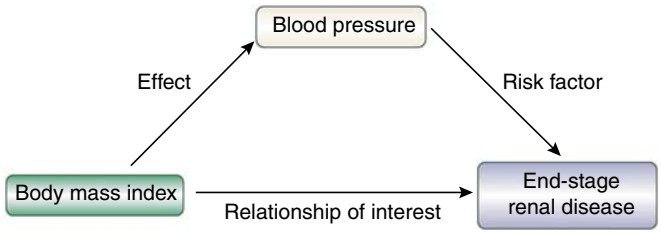

Figure 3 | Relationship between BMI and the development of ESRD—could blood pressure be a confounder in this relationship?

pressure does not satisfy all criteria for a confounder, and in their main analysis, they did not adjust for it.

From these examples, it immediately follows that the analysis of study data itself will not indicate what is or is not a confounder. Only a considerable knowledge of pathophysiological mechanisms and potential causal pathways will assist an investigator to decide whether a variable satisfies the criteria for being a potential confounder or not.

CONTROLLING CONFOUNDING

As confounding obscures the real effect, in other words the etiological importance of a variable, it needs to be prevented or removed as much as possible. Like other types of bias, confounding can be addressed during study design. At that stage, confounding can be prevented by use of randomization, restriction, or matching. In contrast to other types of bias, confounding can also be controlled by adjusting for it after completion of a study using stratification or multivariate analysis. Obviously, adjusting for confounding at this later stage can only take place if information on the confounding factors has been collected during the study.

Randomization

In studies investigating the effects of therapy or other interventions, it is possible to reduce confounding by randomization. As explained in a previous paper in this series,[4] the randomization procedure randomly assigns patients to an experimental group or to a control group. Randomization helps to prevent selection bias by the clinician (sometimes also referred to as 'confounding by indication'). Although randomization of large groups of patients will frequently result in a similar distribution of known and unknown confounders in the experimental and the control group, it is unlikely that this balance will be achieved for all patient characteristics. Although the balance may be incomplete, the randomization process does guarantee that any differences between the two groups are due to chance[5] and not due to the choice of the physician. Thus, although differences in potential confounders between the two groups may still exist after randomization, they are likely to be reduced as much as possible. Examples of large randomized controlled trials in nephrology include the HEMO, ADEMEX, and CREATE studies.[6–8] The size of these randomized controlled trials helped the randomization

process to be successful in producing a similar distribution of known confounders between the experimental groups. The randomized controlled trial performed within the NEtherlands COoperative Study on the Adequacy of Dialysis (NECOSAD),[9] however, was small (only including 38 patients), and therefore, the randomization failed to produce similar experimental groups. The patients randomized to peritoneal dialysis were on an average 7 years younger and had less diabetes mellitus and other comorbidity than those assigned to hemodialysis. As these factors would provide the peritoneal dialysis patients with a better prognosis anyway, the investigators adjusted for these confounders during data analysis to determine the 'real' effect of dialysis modality on outcome.[9]

Other ways of addressing confounding

The other three ways of addressing confounding that will be discussed in this paper are restriction, matching, and stratification. These will be explained by using an example of a hypothetical study investigating the relationship between diabetes mellitus and ischemic heart disease in RRT patients. In such a study, age would be a potential confounder. The simplest way of controlling for confounding by age during the study design would be using *restriction*. The investigators might restrict their study to the group of dialysis patients above the age of 65 years. Although restriction would, at least partially, take away confounding by age, it hampers extrapolation of study results to other groups, in this case to patients below the age of 65 years.

Another method of controlling confounding for age during study design is *matching*. In a cohort study, the patients in the exposed and unexposed groups could be matched in pairs for potential confounders. In the study on the relationship between diabetes and ischemic heart disease, for each 'exposed' person with diabetes mellitus the investigator may select an 'unexposed', that is, nondiabetic, patient of the same age. In this way, the potential confounding effect of age on outcome will be reduced. In cohort studies, the technique of matching is infrequently used, and it may be viewed as a special case of stratification (see later). In case–control studies, however, matching is frequently used. Still, the choice of matching variables needs careful attention because, as will be described later, errors are frequently made.

An approach for controlling confounding after the completion of a study during data analysis is *stratification*. Using this method, the study population in the same example of diabetes and ischemic heart disease is first divided into strata, that is, subgroups according to levels of the potential confounding factor, in this case age. Thereafter, relative risks for each stratum, so-called 'stratum-specific relative risks,' are calculated. Table 1 shows the fictitious data in this study, overall and by age stratum. Suppose that in this study, we found that the crude relative risk is 1.7, meaning that patients with diabetes mellitus are 70% more likely to have ischemic heart disease than patients without diabetes. As age is a potential confounder in this relationship, the study population was divided into strata of patients younger and older than 65 years of age. The relative risk in each stratum was calculated and found to be 1.5 for both age groups. To calculate a summary statistic that describes the effect of diabetes mellitus adjusted for age across the different strata, one may use either pooling (using a Mantel–Haenszel procedure)[10] or standardization. Both methods aggregate information over all strata by taking weighted averages of stratum-specific relative risks to calculate an overall 'adjusted' effect size. Often, one wants to know if a potential confounder is indeed confounding a relationship of interest. If the adjusted effect differs substantially from the crude effect, then confounding is considered to be present. In the example, the age-adjusted effect of diabetes mellitus will be 1.5, as coincidentally, both stratum-specific estimates are 1.5. As 1.7 differs from 1.5, it may be concluded that adjustment for the confounding effect of age is needed to derive an estimate of the real effect of diabetes, as part of the crude effect of 1.7 was found to be due to the effect of age.

Table 1 | Example of confounding in a hypothetical cohort study of ischemic heart disease and diabetes mellitus

	Ischemic heart disease			
	Yes	No	Total	Proportion with ischemic heart disease (%)
All RRT patients[a]				
Diabetes mellitus				
Yes	184	376	560	32.9
No	278	1162	1440	19.3
RRT patients <65 years of age[b]				
Diabetes mellitus				
Yes	36	114	150	24.0
No	136	714	850	16.0
RRT patients ⩾65 years of age[c]				
Diabetes mellitus				
Yes	148	262	410	36.1
No	142	448	590	24.1

RR, relative risk; RRT, renal replacement therapy.
[a]Crude RR=32.9%/19.3%=1.7.
[b]Stratum-specific RR =24.0%/16.0%=1.5.
[c]Stratum-specific RR=36.1%/24.1%=1.5.

Such adjustment for confounding by a particular variable does not always remove all confounding by that variable. There may be residual confounding. For example, a stratified analysis controls confounding only between strata. When there are relatively few strata of continuous variables, such as in the example, where there were only two age strata, there may be substantial residual confounding within each stratum. Stratification can be refined by making more strata, for example, 5-year age bands, and thus will have the effect of improving the adjustment for confounding. Stratification is an effective means for adjusting for confounding when the number of confounding factors is limited. Increasing the number of these factors will rapidly increase the number of strata, as the numbers of categories are multiplied. The

stratification for sex and for the four age categories will use eight strata; further stratification for the presence of diabetes will increase the number of strata to 16. A more practical method of controlling for many confounders at the same time is multivariate analysis. This method will be discussed in future papers in this series.

Finally, it should be noted that adjustment for confounding does not always move the relative risk closer to 1.0, that is, it does not always reduce the differences in outcome between the groups. If the patients in a group with a relative risk of death of 0.8 are older than those in the reference group (by definition 1.0), adjustment for age will increase the difference, by reducing the relative risk in the older group to, for example, 0.7. In summary, adjustment for confounding usually, but not always, reduces the differences in outcome. The next paper in this series on multivariate analysis will provide an example of this.

COMMONLY MADE ERRORS

Available software for statistical analysis has made adjustment for confounding so easy that some investigators have a tendency to 'overadjust' their results. Suppose Hsu *et al.*[3] in Example 2 had not carefully checked beforehand if blood pressure satisfied all criteria for a potential confounder and, like many others in the literature, had simply adjusted for blood pressure to obtain the real effect of BMI. Adjustment for blood pressure would have taken away part of the real effect of BMI, because the effect of an increased BMI may operate via a rise in blood pressure. As a result, the negative effect of a high BMI would have been seriously underestimated. On the other hand, although 'adjustment' for blood pressure is incorrect from the perspective of controlling for confounding, it may be useful to explore potential causal pathways and to generate hypotheses. An example of this can be found in the same paper by Hsu *et al.*[3] These authors showed that additional 'adjustment' for baseline blood pressure levels attenuated the association between BMI and the risk of ESRD: in the main adjusted model without blood pressure, those with a BMI of 35.0–39.9 $kg\,m^{-2}$ had a relative risk of 6.12 (CI: 4.97–7.54) compared to those with a BMI of 18.5–24.9 $kg\,m^{-2}$. After additional 'adjustment' for blood pressure, this relative risk decreased to 4.68 (CI: 3.79–5.79). The degree of change in the effect size of BMI after this adjustment for blood pressure may give the investigator an idea of how much of the effect of an increased BMI acts via the working mechanism of elevated blood pressure.

Similar problems of overadjustment may arise in case–control studies, when cases and controls are matched with respect to a variable that is a potential effect of the exposure. Such might have happened in the study of Janssen *et al.*[11] studying the effect of a polymorphism on the development of diabetic nephropathy in Example 3.

Example 3. CNDP1—Mannheim variant and the susceptibility to diabetic nephropathy *Janssen et al.*[11] *performed a case–control study using diabetic patients with diabetic nephropathy as cases and diabetic patients without diabetic nephropathy as controls. They showed that the CNDP1—Mannheim variant was more common in the absence of diabetic nephropathy (odds ratio 2.56 (CI: 1.36–4.84)).*

An example of incorrect matching would be if these investigators would have decided to match for BMI (or even for glucose intolerance), which are factors that may result from the same polymorphism and may even be in the causal pathway. Most of the real effect of the polymorphism would disappear, and the adjusted odds ratio for the effect of the polymorphism on the development of diabetic nephropathy would be much closer to 1.0.

Both examples show that overadjustment, irrespective of the technique used, may introduce bias instead of preventing it. For this reason, in studies on etiology, the different variables should be carefully checked to determine whether they satisfy all the criteria for a potential confounder. Second, in case–control studies, the matching procedure itself (that is intended to reduce confounding) can introduce confounding, even when the matching variable satisfies all criteria for a confounder. This occurs because, by matching for this confounder, the cases and controls will start to appear more alike, not only for this specific confounder, but also for other variables that are related to this confounder. For example, matching for age and sex will also make cases and controls more alike with respect to BMI. If the confounder that is used as a matching variable is strongly related to the risk factor causing the outcome, matching will lead to an underestimation of the effect of that risk factor. Therefore, in case–control studies, matching for confounding may result in overadjustment and even introduce confounding. To remove this 'new' confounding, adjustment for the matching variables is necessary.[12]

Another commonly made mistake arises from the use of tests of statistical significance to detect confounding.[13] The amount of confounding, however, is the result of the strength of the associations between the confounder on the one hand and the exposure and the disease on the other hand. *P*-values will therefore not provide information if a particular variable is a confounder. As explained in one of the previous paragraphs, the amount of confounding caused by a variable that satisfies all criteria for a potential confounder can be measured by looking at the difference between the crude and adjusted effect size. If these are almost equal, there is no confounding, but if the difference between the two is important, there is confounding.

CONCLUSION

Confounding in etiological studies can be described as a 'mixing' of effects distorting the real effect of an exposure. As a result, a crude effect may not equal the 'true' effect of a risk factor. Before adjusting for confounding, the criteria for a possible confounder should be carefully checked to prevent the introduction of new bias through overadjustment for variables that do not satisfy all criteria for confounding.

REFERENCES

1. Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet* 2002; **359**: 248–252.
2. Couchoud C, Moranne O, Frimat L *et al.* Associations between comorbidities, treatment choice and outcome in the elderly with end-stage renal disease. *Nephrol Dial Transplant* 2007 e-pub ahead of print 5 July 2007.
3. Hsu C, McCulloch CE, Iribarren C *et al.* Body mass index and risk for end-stage renal disease. *Ann Intern Med* 2006; **144**: 21–28.
4. Stel VS, Jager KJ, Zoccali C *et al.* The randomized clinical trial: an unbeatable standard in clinical research? *Kidney Int* **72**: 539–542.
5. Vandenbroucke JP. When are observational studies as credible as randomized trials? *Lancet* 2004; **363**: 1728–1731.
6. Eknoyan G, Beck GJB, Cheung AK *et al.* Effect of dialysis dose and membrane flux in maintenance hemodialysis. *N Engl J Med* 2002; **347**: 2010–2019.
7. Paniagua R, Amato D, Vonesh E *et al.* Effects of increased peritoneal clearances on mortality rates in peritoneal dialysis: ADEMEX, a prospective, randomized, controlled trial. *J Am Soc Nephrol* 2002; **13**: 1307–1320.
8. Drüeke TB, Locatelli F, Clyne N *et al.* Normalization of hemoglobin level in patients with chronic kidney disease and anemia. *N Engl J Med* 2006; **355**: 2071–2084.
9. Korevaar JC, Feith GW, Dekker FW *et al.* Effect of starting with hemodialysis compared with peritoneal dialysis in patients new on dialysis treatment: a randomized controlled trial. *Kidney Int* 2003; **64**: 2222–2228.
10. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959; **22**: 719–748.
11. Janssen B, Hohenadel D, Brinkkoetter P *et al.* Carnosine as a protective factor in diabetic nephropathy. Association with a leucine repeat of the carnosinase gene CNDP1. *Diabetes* 2005; **54**: 2320–2327.
12. Rothman KJ, Greenland S. *Modern Epidemiology*. 2nd edn. Lippincott: Philadelphia, 1998.
13. Rothman KJ. *Epidemiology: An Introduction*. Oxford University Press: New York, 2002.