Journal of Clinical Epidemiology 121 (2020) 91−100

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

A nontechnical explanation of the counterfactual definition of confounding

Martijn J.L. Bours*

Department of Epidemiology, GROW − School for Oncology and Developmental Biology, Maastricht University, Maastricht, the Netherlands

Accepted 24 January 2020; Published online 14 February 2020

Abstract

In research addressing causal questions about relations between exposures and outcomes, confounding is an issue when effects of interrelated exposures on an outcome are confused. For making valid inferences about cause-and-effect relationships, the biasing influence of confounding must be controlled by design or eliminated during data analysis. Consequently, researchers require a sound understanding of the concept of confounding to adequately deal with this type of bias when setting up and conducting (clinical) epidemiological research. For explaining confounding on a conceptual level, the counterfactual framework for causal inference is invaluable but can be very complicated. In this article, therefore, a nontechnical explanation of the counterfactual definition of confounding is presented. When considering confounding in a counterfactual way, the principle of exchangeability plays a pivotal role. Causal effects of an exposure on an outcome can be evaluated only when different exposure groups have comparable background risks of the outcome. Then, exposure groups are exchangeable and thus unconfounded. By providing a simplified explanation of the counterfactual principles of exchangeability, and consequences of nonexchangeability, this article aims to increase understanding of confounding on a conceptual level as well as the rationale underlying design and analytic strategies for dealing with confounding in (clinical) epidemiological research.

Keywords: Confounding; Bias; Counterfactual theory; Exchangeability; Causality

1. Confounding biases the process of causal inference about a target population

Confounding may well be one of the most frequently discussed concepts in the (clinical) epidemiological literature. When confronted with the concept of confounding for the first time, and perhaps also thereafter, it may be difficult to fully grasp its meaning and underlying principles. A first step toward understanding confounding is a clear definition. According to the Dictionary of Epidemiology [1], confounding is defined as follows:

> *The distortion of a measure of the effect of an exposure on an outcome due to the association of the exposure with other factors that influence the occurrence of the outcome. Confounding occurs when all or part of the apparent association between the exposure and the outcome is in fact accounted for by other variables that affect the outcome and are themselves not affected by the exposure.*

Central players in the definition of confounding are exposures and outcomes and, particularly, their potential causal relationship on which much (clinical) epidemiological research is focused. In such research, an exposure is broadly defined as being exposed to some kind of determinant, either harmful (risk factor) or beneficial (protective factor), or to a certain intervention or treatment. Like exposures, outcomes of interest in (clinical) epidemiological research are also broadly defined; for example, the occurrence or cure of a certain disease or health-related condition. In this article, unless stated otherwise, it is assumed for ease of explanation that exposures and outcomes are dichotomous and are positively related, meaning for instance that exposure to a risk factor leads to more disease or exposure to a treatment leads to more cure of disease. However, the concepts to be explained also apply to exposures and outcomes that are nondichotomous or inversely related.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest: The author is a member of the editorial board of the Journal of Clinical Epidemiology but was not involved in the peer review process or decision to publish.

* Corresponding author. Department of Epidemiology, Maastricht University, P.O. Box 616, 6200 MD Maastricht, the Netherlands. Tel.: +31 0 433882903; fax: +31 0 433884128.

E-mail address: m.bours@maastrichtuniversity.nl.

https://doi.org/10.1016/j.jclinepi.2020.01.021

What is new?

Key findings

- Confounding threatens the process of causal inference in (clinical) epidemiological research, yet its underlying principles are complex and often misunderstood.
- A sound understanding of confounding within the counterfactual framework of causation enables better anticipation and dealing with this source of bias in research practice.

What this adds to what was known?

- A simplified explanation of the counterfactual definition of confounding is provided, based on a nontechnical and graphical presentation of the central role of exchangeable background risks.

What is the implication and what should change now?

- This article can be used as a teaching tool for introducing researchers (and students) to the underlying concepts of confounding explained from a counterfactual viewpoint.

When conducting experimental and, especially, observational studies to address causal questions about exposure-outcome relations in a certain target population, researchers must always be on the lookout for the threat of confounding that may bias causal inferences about the target population under study [2]. Because the target population is defined by the particular causal research question of interest, confounding is a target population-specific concept. The co-occurrence of multiple exposures and their intertwined influence on an outcome within a specific target population is an actual (biological) fact and in itself no bias. As a result of existing interrelations between multiple exposures, however, it is not easy to disentangle the effect of the primary exposure of interest (e.g., a risk factor or treatment) from the effect of one or more other exposures that confound the exposure-outcome relationship in the target population under investigation. Consequently, when not appropriately accounted for by design or in the analysis, confounding biases study findings by distorting the association measures used for quantifying the nature and magnitude of the relation between the primary exposure and outcome [3−8]. It is a principal task of the researcher to attempt to eliminate this bias or at the very least try to reduce it, by applying either appropriate design or analytic strategies (or both).

In the literature, several definitions of confounding and characteristics of confounders have been put forward [8,9]. Traditionally, it was proposed that an extraneous variable confounds an exposure-outcome association when meeting three criteria [10,11]. The potential confounder must be related to both the exposure and outcome of interest but must be affected by neither of them. Although these traditional criteria clarify essential properties of a confounding variable, they are considered inadequate because looking for evidence of the separate criteria can be misleading if one relies only on observed relations in study data without properly minding the causal context within the target population [9,12]. Therefore, relevant subject-matter knowledge should always be used when identifying potential confounders of the exposure-outcome relation(s) under study. A useful tool for this purpose are causal diagrams, also called directed acyclic graphs [13,14].

Causal diagrams graphically depict theory-driven assumptions about causal relations between the exposure and outcome variables in a target population under study and any other variables directly or indirectly related to the study variables. Correctly specified causal diagrams based on appropriate expert knowledge constitute a strong tool for identifying potential confounding variables and for distinguishing confounders from nonconfounders. Fig. 1 shows how a confounder is defined and can be identified in a causal diagram. Furthermore, Fig. 1 also shows how causal diagrams can be used to define other types of variables that should be distinguished from confounders (i.e., mediators and colliders), which is a particular strength of using carefully constructed diagrams. The concepts of mediator and collider variables are beyond the scope of this article, however, and will therefore not be explained further. Confounding can be introduced by a single variable or set of variables, which should be minimally taken into account before drawing causal conclusions about an exposure-outcome relation of interest. In that sense, a confounder can be defined as a variable for which control by design or analytical adjustment is required to obtain unbiased estimates of the effect of an exposure on the outcome under study [9]. Basically, confounding thus reflects a special kind of triangular relationship between an outcome and two (or more) interrelated exposures that both affect the outcome. Which of the exposures is regarded as the primary exposure, whose unbiased relation with the outcome is of interest, and which one is treated as the confounder, whose biasing influence needs to be eliminated, depends mostly on the main research questions and causal structures that have been hypothesized a priori [15].

It is extremely important for researchers addressing causal questions to master the principles of confounding because it will enable them to better anticipate and deal with this common type of bias. One way to gain a deeper understanding of the concept of confounding is through the counterfactual theory of causation [16−20]. Counterfactual theory has gained popularity as a way to define and statistically quantify cause-and-effect relations, as well as the types of bias, including confounding, that threaten the interpretation of these relations. Basic knowledge about counterfactuals can help better understand how

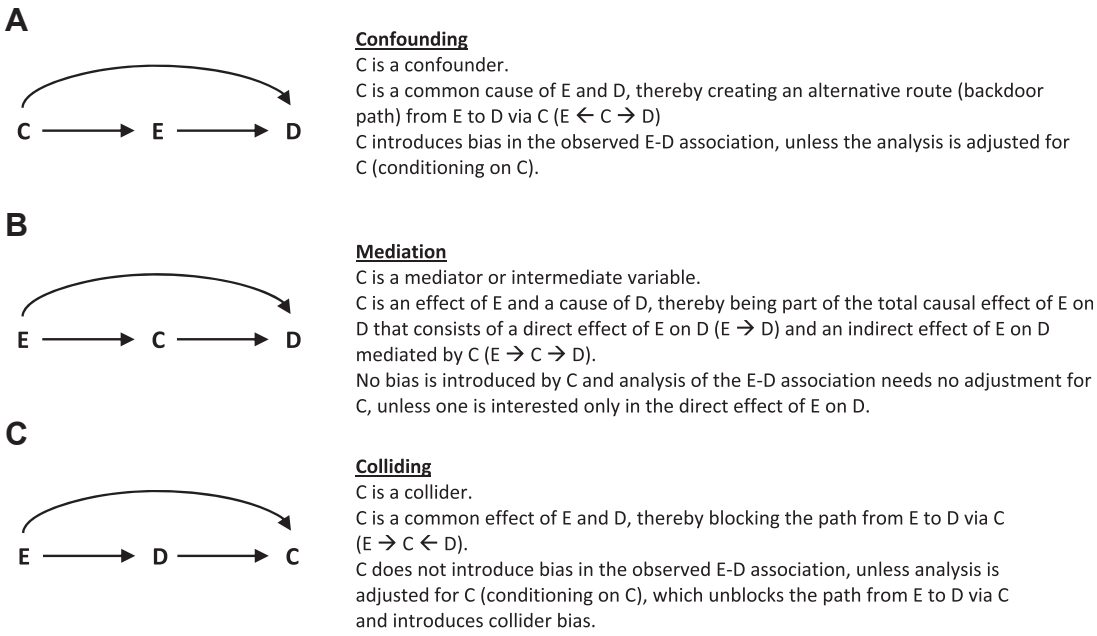


Fig. 1. Causal diagrams showing possible triangular relationships between an exposure E, an outcome D, and a third variable C. Confounding of the causal E-D relation occurs when variable C is a cause of both exposure E and outcome D (A). In case variable C is a cause of outcome D but an effect of exposure E, then C is not a confounder but a mediator (B). If variable C is an effect of both exposure E and outcome D, then C is a collider instead of a confounder (C).

confounding can bias the process of causal inference. Therefore, the aim of this article is to present a nontechnical explanation of the definition of confounding within the counterfactual framework.

2. Understanding the counterfactual framework increases understanding of confounding

A starting point for explaining confounding on a conceptual level is imagining a situation in which it cannot occur. Causal reasoning in hypothetical contrasts (counterfactuals) can be very insightful for that purpose and thereby useful for conceptually defining confounding [3,7,16–18]. In simple terms, the counterfactual theory of causation could be called the 'What if?' theory. When thinking in a counterfactual way about a causal, unconfounded effect of an exposure on an outcome, the potential outcomes in two situations with contrasting states of exposure are compared (e.g., exposed vs. unexposed). One of these situations is real (factual) and can potentially be observed, whereas the other is unreal (counterfactual) and cannot be observed. If for instance the actual observed situation is the exposed state, then the sixty-four thousand dollar question about causation is 'What would the outcome have been if exposed persons had been unexposed?' Accordingly, an exposure has a causal effect on an outcome in a specific target population if and only if the occurrence of the outcome in exposed persons is different from the occurrence in the same persons, when they would not have been exposed. As only one of these two potential outcomes can obviously be observed in reality, the theory of counterfactuals has been criticized for lacking practical applicability [21,22]. However, in recent decades, the counterfactual way of thinking has enabled the statistical quantification of causal effects and has fueled, among other things, the development of causal graph theory [23,24]. Therefore, understanding how causal effects are defined according to counterfactual theory facilitates understanding of what noncausal effects due to confounding are.

2.1. A counterfactual scenario precludes confounding: central role of the background risk

To answer the question how the concept of *no* confounding is explained based on counterfactuals, one could imagine a study scenario that precludes confounding. As a thought experiment, suppose you perform a study on the effect of some exposure (e.g., a certain treatment) on the occurrence of an outcome of interest, for instance the recovery from a certain disease. Imagine also living in the hypothetical world of H.G. Wells and so having a time machine at your disposal. You could then perform a true counterfactual study. First, you expose a group of diseased individuals (group A) to the treatment and, after some follow-up time, determine the incidence of recovery from disease as a measure of the risk of the outcome of interest. Next, you travel back in time using your time machine and select the same group of diseased individuals again (group A′). Now, however, you do *not* expose them to the treatment and, under the exact same conditions and after the same amount of follow-up time, you determine the incidence of recovery once more. Based on this counterfactual

study, assuming the treatment increases the 'risk' of recovery, you could determine two risks in one and the same group of individuals: (1) the risk in the unexposed group A′ that is independent of the treatment (the background risk), and (2) the risk in the exposed group A, which is the background risk *plus* the extra risk due to the treatment.

Comprehension of the meaning of the background risk is crucial here. This risk is determined by person-, time-, and place-related factors that constitute the causal context for an individual or group of individuals. It is the risk of the outcome of interest that is not related to the primary exposure but caused by other exposures that are specific to the individuals being studied. In the counterfactual study, any absolute or relative difference in the 'risks' of recovery between the exposed and unexposed situations has to be caused by exposure to the treatment because all other factors related to person, time, and place were identical in these two situations. Consequently, confounding by extraneous variables is impossible in the counterfactual study since no factors whatsoever differed between the exposure groups except for the treatment under investigation. In other words, the causal contexts were identical in group A and A′, which after all was one and the same group. Any measure of effect, such as the risk difference (RD), risk ratio (RR), or odds ratio (OR), will thus be free of confounding and reflect the causal effect of the exposure on the outcome, both at the level of the individual and the group of individuals. Fig. 2 depicts the counterfactual situation of no confounding. In addition, for a better understanding of how causal effects at the individual and at the population level are defined according to counterfactual theory, a definition of causal subtypes and how this relates to the concept of the background risk is provided in a web-only appendix.

2.2. No confounding in research practice in case of exchangeable background risks

Of course, in reality, one can never study an exposure-outcome relation in the same group of individuals under identical conditions at exactly the same time and place. In the practice of (clinical) epidemiological research, groups of different individuals are compared or groups of the same individuals are compared at different times. An observed association between an exposure and outcome of interest based on a comparison of distinct study groups is an estimate of the causal effect of the exposure on the outcome in the target population of interest [5,6,17,18]. Because the primary exposure is not the only determinant of the outcome of interest that can vary under such circumstances, the possibility of confounding by exposure to these other determinants is introduced. In research practice, one must therefore endeavor to mimic the counterfactual situation as much as possible by comparing exposure groups comprised of individuals who have comparable background risks of the outcome. This means that when studying for instance two groups, exposed vs. unexposed, the unexposed group should represent the counterfactual situation for the exposed group if this group had not been exposed, and vice versa. Ideally, the actual exposure groups thus need to be valid substitutes of the counterfactual exposure groups [17,18]. Otherwise, one would not only be estimating the effect of the primary exposure on the outcome but also effects of other exposures that confound the exposure-outcome relation under study, hampering the ability to draw causal conclusions.

As an example of how the counterfactual situation of no confounding could be mimicked in research practice, suppose you perform a randomized study with two groups of

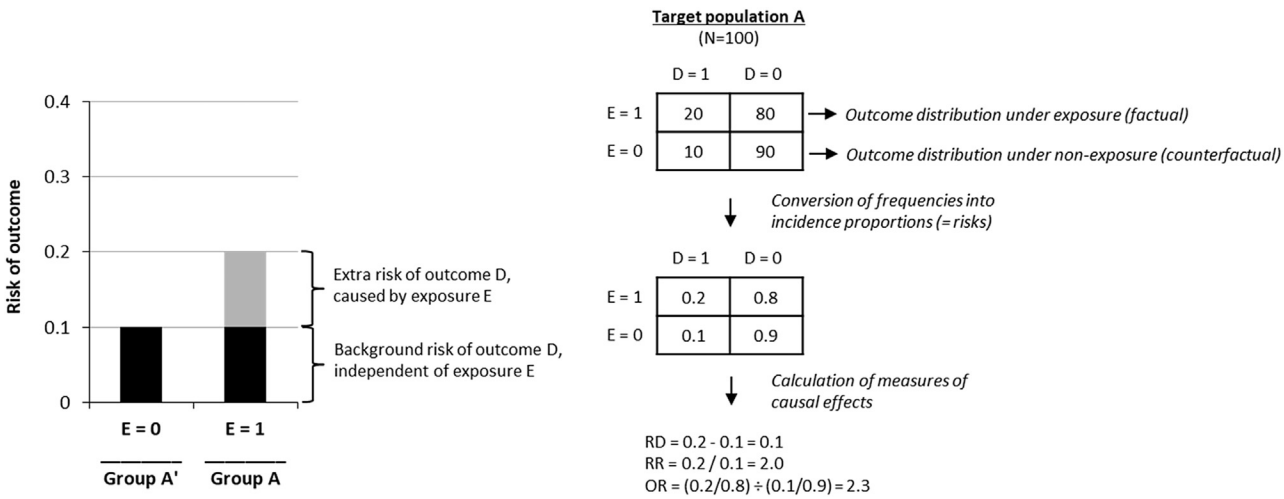


Fig. 2. The counterfactual situation of no confounding. The effect of an exposure E on the risk of outcome D in a single target population A is shown in two contrasting conditions: exposed (A) and unexposed (A′). The absolute or relative difference in risk of outcome D between the two conditions indicates the true (unconfounded) effect of exposure E on D, as expressed by causal measures of effect. (E = 1 and E = 0 indicate presence and absence of exposure, respectively, and D = 1 and D = 0 indicate presence and absence of the outcome, respectively; RD, risk difference; RR, risk ratio; OR, odds ratio).

different individuals, one of which is exposed (group B) and one which is not exposed (group C) to some treatment. Assume that this study is performed in an infinitely large population so that the probability of random imbalances between the study groups approaches zero. After a certain period of follow-up, the frequency (risk) of the occurrence of some outcome is determined in both randomly allocated groups. Just as in the above-described counterfactual study, two risks can now be distinguished: (1) the background risk (group C), and (2) the background risk plus the extra risk due to the treatment (group B). If the average background risk of the outcome is the same in both study groups at baseline (and remains the same during follow-up), any absolute or relative risk difference between the two groups results from the effect that being exposed to the treatment had on the outcome. Consequently, any measure of association will not be biased by confounding (Fig. 3).

The crucial aspect in this example of no confounding is again the background risk of the outcome that is not caused by, and thus independent of, the exposure of interest [5–7,17,18]. When comparing exposure groups of different individuals, the background risk at the level of the individuals is unlikely to be identical because different individuals likely have different causal contexts and may represent different causal subtypes [8,17,18]. However, if the average background risks are comparable between exposure groups of different individuals, association measures based on a comparison of these groups will be valid estimates of the average causal effect that the exposure has on the outcome. In case of comparable background risks, exposure groups are said to be *exchangeable* [3–5,7,18]. This means that if, by the flip of a coin, the treatment allocation in the randomized study had turned out to be the other way around (i.e., group C instead of B exposed to the treatment), the exact same measures of association would have been observed (Fig. 3). To accomplish exchangeability of exposure groups, the most effective strategy is randomization [3,7,18,25]. By randomly allocating individuals into one group exposed and another group not exposed to some treatment, the likelihood of the allocated groups being exchangeable with regard to their background risks is increased parallel to an increase in the sample size. The larger the sample size, the larger the chance that the randomization process balances the distribution of potential confounders in different exposure groups [26], thereby creating exchangeability of background risks. Another way to increase the likelihood of exchangeability is through restriction of potential confounders, as a result of which only individuals with certain predefined characteristics are included in a study [3,7,23]. Homogeneous exposure groups are created in this way to reduce the chance of confounding by the restricted variables as well as correlates thereof.

2.3. *Confounding occurs if background risks are nonexchangeable between exposure groups*

So when does confounding occur? As might be clear from the aforementioned explanations of no confounding, an exposure-outcome association will be confounded when the background risks of the outcome of interest are dissimilar between the exposure groups to be compared [3–5,7,8,18]. Exposure groups are then not exchangeable in a counterfactual sense, leading to a confounded, 'apples-and-oranges' comparison. Another example will clarify how nonexchangeability of background risks can confound associations between an exposure and an

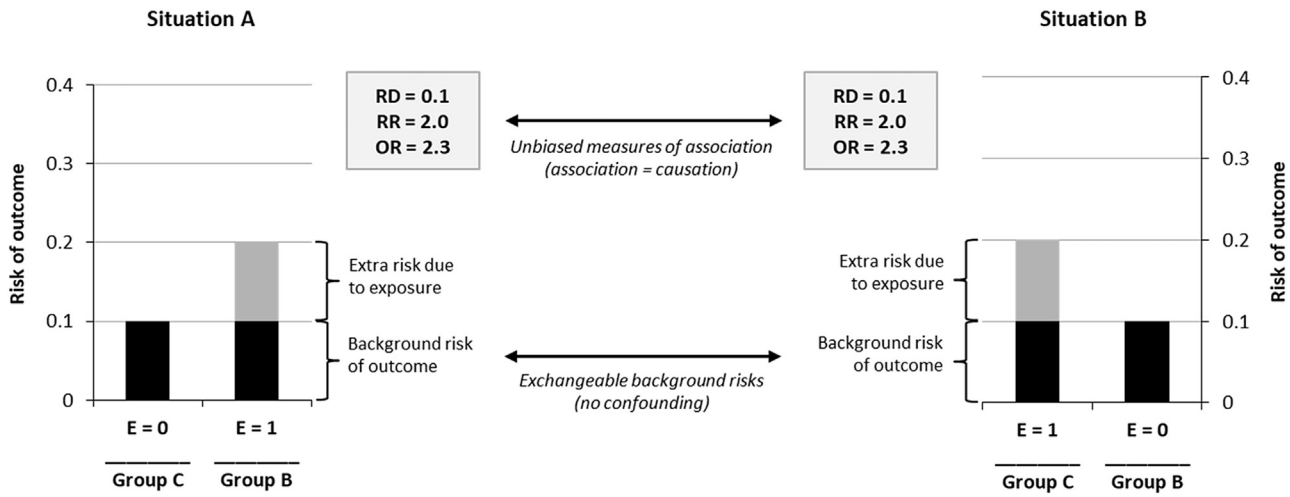


Fig. 3. The situation of no confounding in research practice. The effect of an exposure E on the risk of an outcome is studied by comparison of two different groups (B and C). Two contrasting situations are depicted: group B exposed and group C unexposed (situation A) vs. group C exposed and group B unexposed (situation B). The identical measures of association in the two situations indicate that exchangeability of background risks is a prerequisite for observing unconfounded associations, showing that the causal effect of exposure E is identifiable in the absence of confounding (association = causation). (E = 1 and E = 0 indicate presence and absence of exposure, respectively; RD, risk difference; RR, risk ratio; OR, odds ratio).

outcome in the practice of (clinical) epidemiological research.

Once more, imagine performing a study. You now conduct an observational study on the association between exposure to some risk factor and the occurrence of some disease. A cohort of individuals at risk of the disease is enrolled and, after a certain follow-up period, the incidence of the disease in both the exposed subcohort (group D) and the unexposed subcohort (group E) is determined. If the average background risk of the disease differs between the exposed subcohort and the unexposed subcohort, any measure of association between the risk factor and the disease based on a comparison of disease incidences will be confounded by the factor(s) responsible for the different background risks (Fig. 4). Confounding occurs because when being exposed or not is related to other disease determinants, the background risks in the subcohorts will not be exchangeable anymore. This nonexchangeability means that if exposure to the risk factor had been the other way around for some reason (group E instead of D exposed), the measures of association would change, indicating that causal effects are not identifiable in the presence of confounding (Fig. 4).

3. The nature of nonexchangeable background risks determines the direction of confounding bias

Confounding due to nonexchangeable background risks between exposure groups can bias measures of an exposure-outcome relation in different directions. When the average background risk of the outcome of interest is lower among exposed persons than among unexposed persons, the bias in the observed association between exposure and outcome will be downward, resulting in underestimation of the causal effect that the exposure has on the outcome (situation A in Fig. 4). By contrast, if the background risk among exposed persons is higher than among unexposed persons, the observed association will be biased upward, resulting in overestimation of the causal effect of the exposure (situation B in Fig. 4). Note that the above-described direction of the bias holds true only for an exposure that increases the risk of an outcome. In case an exposure decreases the risk of an outcome, the direction of the bias resulting from nonexchangeable background risks between exposure groups will be the opposite. Moreover, if the nonexchangeability becomes too large, a true causal relation between an exposure and outcome may be completely masked or even reversed. The latter case of extreme confounding would result in bias across the null (e.g., a true risk factor may spuriously appear as a protective factor). To illustrate potential consequences of nonexchangeability in the practice of (clinical) epidemiological studies, two examples of confounding are presented in Box 1.

4. Design and analytic strategies to combat confounding are based on nonexchangeability principles

Confounding is always a possibility to be considered as an alternative, noncausal explanation of observed exposure-outcome associations in experimental and, especially, observational studies. Researchers have the crucial task to try to control confounding at the design stage of a study and/or attempt to eliminate it at the data analysis stage. As already mentioned, common design strategies to control for confounding are randomization and restriction, which aim to enhance exchangeability between exposure groups.

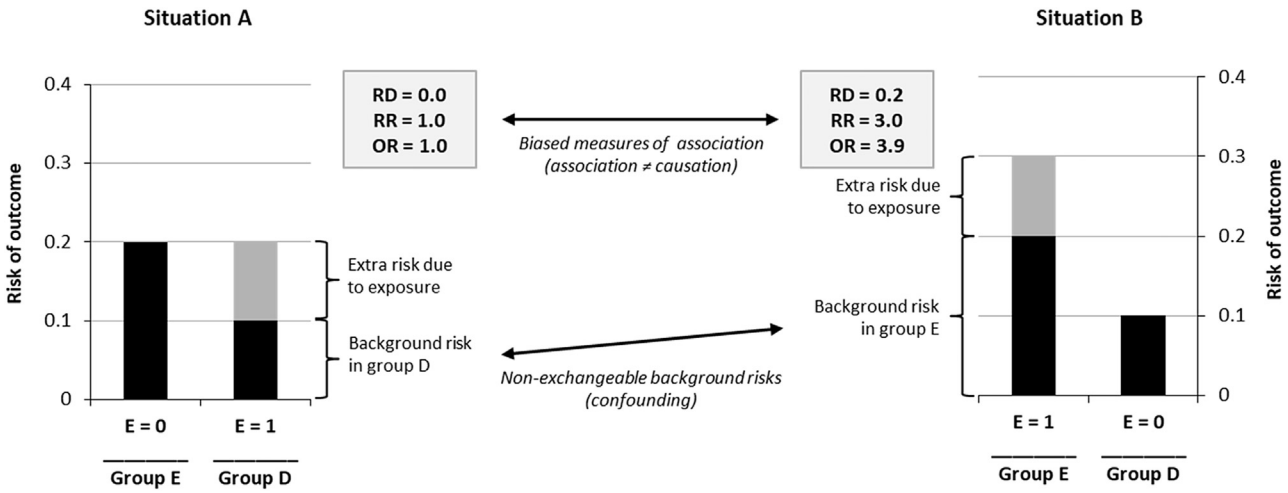
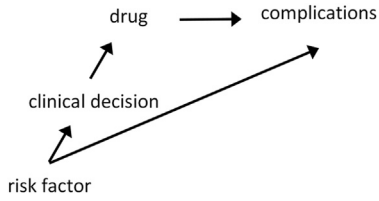


Fig. 4. The situation of confounding in research practice. The effect of an exposure E on the risk of an outcome is studied by comparison of two different groups (D and E). Two contrasting situations are depicted: group D exposed and group E unexposed (situation A) vs. group E exposed and group D unexposed (situation B). The dissimilar measures of association in the two situations indicate that nonexchangeability of background risks induces confounding of observed associations, showing that the causal effect of exposure E is not identifiable in the presence of confounding (association ≠ causation). Association measures in both situations are confounded by factors responsible for the difference in background risks between exposure groups. (E = 1 and E = 0 indicate presence and absence of exposure, respectively; RD, risk difference; RR, risk ratio; OR, odds ratio).

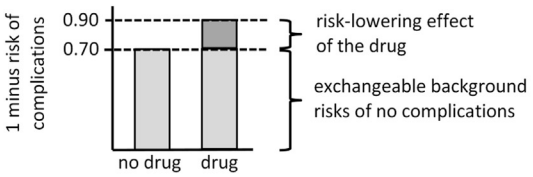
Box 1 Practical examples of confounding

Confounding in experimental research: confounding by indication

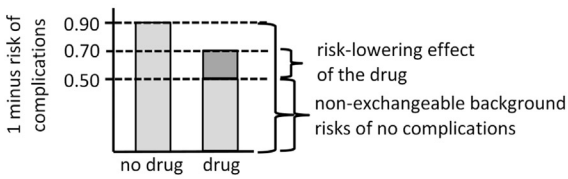
In nonrandomized controlled trials, or even in randomized controlled trials in which the random treatment allocation is not adequately concealed, confounding can occur if the decision to allocate the treatment under investigation depends on a factor that is also related to the outcome of interest. Imagine for instance an experimental study in which the effect of anti-inflammatory medication (e.g., a drug) on the occurrence of postoperative complications (e.g., wound infection) is evaluated. The surgeon decides who receives the drug after surgery and who does not. If the surgeon's decision to prescribe the drug is influenced by specific patient characteristics, such as a risk factor (e.g., advanced age) that is a clinical indication for prescribing the drug to lower the risk of postoperative complications (see causal diagram), then confounding by indication will bias the effect of the drug on postoperative complications.



Suppose that the true effect of the drug is to lower the risk of complications by 20% and that the average background risk of complications is 30% in the total study population. Further suppose that because the surgeon prescribed the drug more often to patients with the risk factor than to patients without, the background risk of complications is 50% in the patients receiving the drug and 10% in the patients not receiving the drug. The bar graphs show the situation of no confounding (left) and the situation of confounding (right) due to the nonexchangeability of background risks induced by the confounding by indication. Assuming that the true risk-lowering effect of the drug remains unchanged and thus is independent of the risk factor responsible for the difference in background risks, the confounding by indication has produced bias across the null (i.e., reversal of effect), leading to the false conclusion that the drug has resulted in an increased risk of complications.



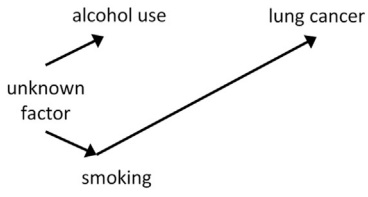
Unconfounded measures of the effect of medication:
RD = 0.10 – 0.30 = -0.20
RR = 0.10 / 0.30 = 0.33
OR = (0.10/0.90) ÷ (0.30/0.70) = 0.26



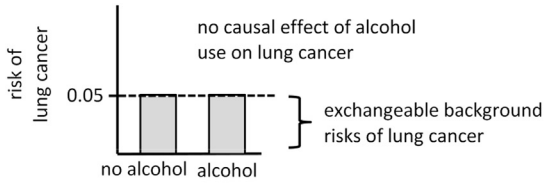
Confounded measures of the effect of medication:
RD = 0.30 – 0.10 = 0.20
RR = 0.30 / 0.10 = 3.00
OR = (0.30/0.70) ÷ (0.10/0.90) = 3.86

Confounding in observational research: spurious association induced by confounding

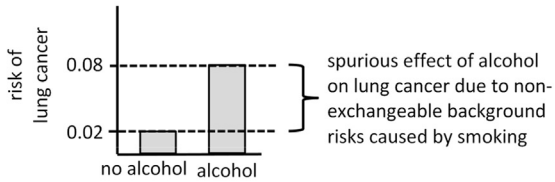
Confounding can induce spurious (noncausal) associations between an exposure and outcome that are not causally related but share a common cause. Suppose that the relation between alcohol and lung cancer is studied within a prospective cohort study. Alcohol use is not causally related to lung cancer, but smoking is a strong risk factor for developing lung cancer. Because alcohol use and smoking are associated, probably linked by some unknown factor (e.g., a certain personality trait) that causes both alcohol use and smoking (see causal diagram), an association between alcohol use and lung cancer can be observed because of confounding by smoking.



Suppose that the average background risk of lung cancer in the study population is 5% and that in reality alcohol use does not causally affect the risk of lung cancer. This means that the risk of lung cancer in alcohol users and nonusers is equal to the background risk, which is determined by factors other than alcohol use, including smoking. Further suppose that smoking is positively related to alcohol use (i.e., alcohol users are more likely to also smoke compared to nonusers) and that smoking increases the risk of lung cancer. The bar graphs show the situation of no confounding (left) and the situation of confounding (right) due to nonexchangeability of background risks resulting from confounding by smoking. Assuming that alcohol has no causal effect on lung cancer in the absence and in the presence of smoking (no interaction between alcohol and smoking), the true measures of effect are null. However, if the confounding by smoking is disregarded, the background risk of lung cancer in alcohol users is increased relative to nonusers because the former group contains more smokers. As a result, the confounding by smoking then produces a spurious association between alcohol and lung cancer that does not exist in reality, leading to the false conclusion that alcohol use increases the risk of lung cancer.



Unconfounded measures of the effect of alcohol use:
RD = 0.05 – 0.05 = 0.00
RR = 0.05 / 0.05 = 1.00
OR = (0.05/0.95) ÷ (0.05/0.95) = 1.00



Confounded measures of the effect of alcohol use:
RD = 0.08 – 0.02 = 0.06
RR = 0.08 / 0.02 = 4.00
OR = (0.08/0.92) ÷ (0.02/0.98) = 4.26

In effect, analytic strategies to deal with confounding are also based on exchangeability principles, including multivariable regression modeling, stratification, standardization, inverse probability weighting, and propensity score methods [23]. How does analytical adjustment for confounding work from a counterfactual perspective? Take, for example, a stratified analysis, which is a simple method to adjust for an imbalanced distribution of confounders between exposure groups. The goal of stratification is to create homogeneous subgroups (strata) within the study population, in which confounder distributions are balanced between exposure groups. Stratification eliminates the relation between the confounder and exposure of interest, thereby ensuring conditional exchangeability of background risks within confounder strata (Fig. 5). Consequently, differences in risks of the outcome between stratified exposure groups are unconfounded and can thus be attributed to the exposure of interest. The same principle basically applies to the other analytic strategies mentioned above, which aim to create conditional exchangeability through conditioning on one or more confounders. It is important to note that when applying analytical methods to adjust for confounding, issues such as incomplete adjustment (residual confounding) or adjusting for mediators (overadjustment bias) [27] must be avoided as much as possible and should always be considered when drawing inferences about unconfounded causal effects of an exposure on an outcome in a target population of interest.

Of note, next to dealing with the lurking threat of confounding by mixing up effects of two or more exposures, researchers must also consider the possibility of interplay between effects of exposures on some outcome of interest [28−32]. This refers to the concepts of effect modification and interaction, which are related to the concept of confounding and can complicate interpretation of data analyses because of scale dependency. In fact, as can be noticed from the example of the stratified analysis shown in Fig. 5, the relative association measures (RR and OR) were not uniform across confounder strata, whereas the absolute association measure (RD) was. The heterogeneity of the exposure-outcome association when expressed on the risk-ratio or odds-ratio scale may indicate that the effects of the exposure and the stratification variable on the outcome are interdependent because it appears that the exposure effect is somehow modified by the stratification. A relevant question is whether this represents causal effect modification or interaction, that is, whether the causal effect of the exposure is truly changed under the influence of the stratification variable. Although in principle this could be a possibility, as a confounding variable can also be a modifier of the effect that an exposure has on an outcome, the answer is no in the particular example shown in Fig. 5 because the effect of the exposure on the risk-difference scale is homogeneous within the strata. The magnitude of the absolute exposure effect thus remains unchanged, but because the magnitude of the background risks differs between the strata, the stratified relative exposure effects also differ as they incorporate the background risk. This algebraic phenomenon is the reason why effect modification is sometimes called effect-measure modification to refer to the scale-dependency that complicates the

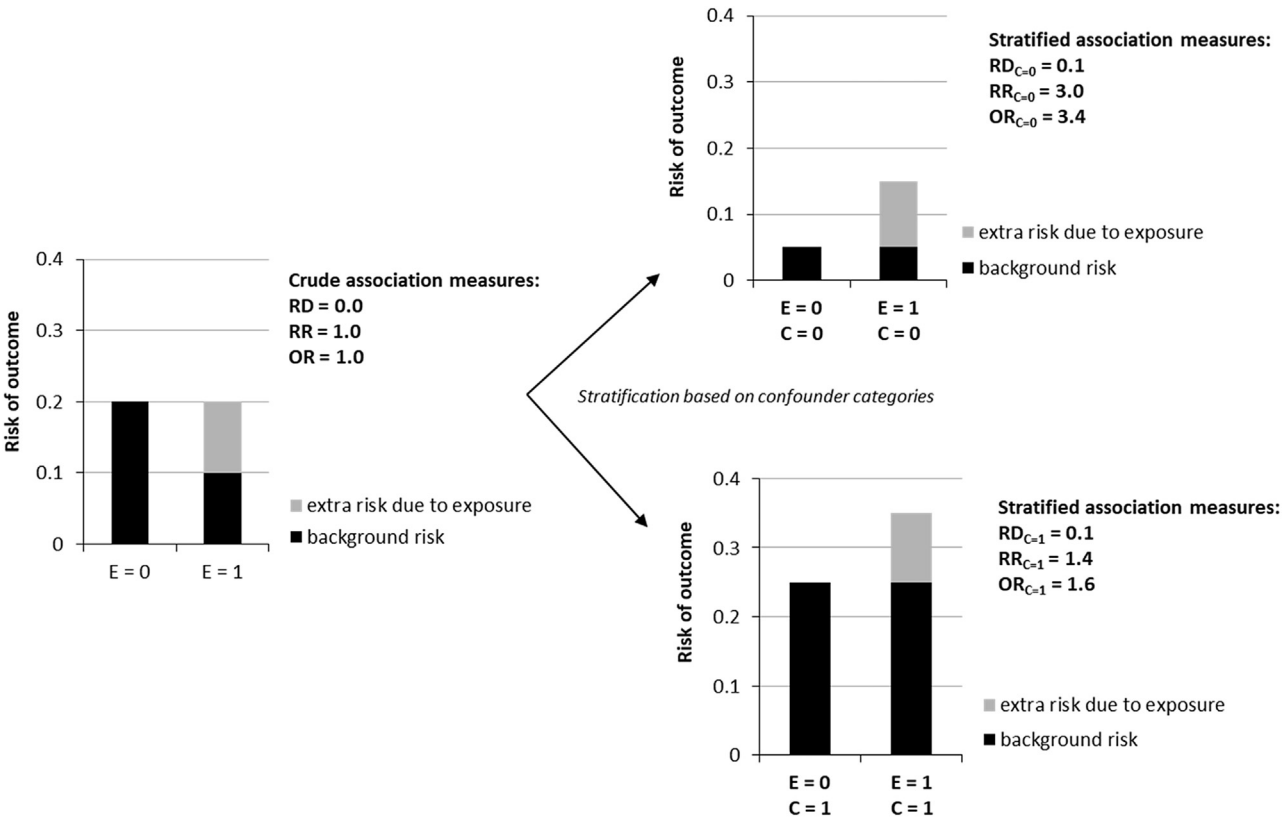


Fig. 5. Example of a stratified analysis to account for confounding of crude measures of association between exposure E and an outcome due to nonexchangeability of background risks caused by variable C. Stratification creates conditional exchangeability of background risks between exposure groups within strata based on categories of confounder C. As a result, stratified association measures are unconfounded, reflecting that the causal effect of exposure E is identifiable after conditioning on confounder C. (E = 1 and E = 0 indicate presence and absence of exposure, respectively, and C = 1 and C = 0 indicate the presence and absence of the confounder, respectively; RD, risk difference; RR, risk ratio; OR, odds ratio).

evaluation and interpretation of effect modification or interaction [32,33]. For example, in case of a stratified analysis, if stratified association measures are uniform on an absolute (additive) scale, they will automatically not be on a relative (multiplicative) scale, and vice versa. Measures of association cannot be homogeneous on both scales, except for the situation when the exposure of interest is not causally related to the outcome and all measures of effect are truly null.

Finally, when evaluating confounding based on a comparison of crude (unadjusted) and stratified (adjusted) effect estimates, an important issue to consider is the crucial distinction between so-called collapsible effect measures such as the RD and RR and noncollapsible measures such as the OR. Caution is warranted if the OR is used to evaluate confounding because the definition of confounding based on noncollapsibility principles (i.e., confounding in case crude and adjusted effect estimates differ meaningfully) does not necessarily converge with the definition of confounding based on exchangeability principles (i.e., confounding in case of nonexchangeable background risks between exposure groups) [19]. Indeed, under certain specific conditions, described in detail elsewhere, noncollapsibility of the OR can occur in the absence of confounding and collapsibility of the OR can occur in the presence of confounding [5−7,10]. It is therefore generally not recommended to use noncollapsible effect measures to evaluate confounding based on a comparison of crude and adjusted effect estimates, as any nonrandom difference can be the result of confounding, noncollapsibility, or both.

5. Summary

Researchers performing (clinical) epidemiological studies must anticipate and eliminate confounding, either through control by design or adjustment during data analysis. At the very least, the influence that confounding could have had on observed study findings needs to be considered in the process of drawing causal inferences about observed associations between some exposure and outcome in a target population of interest. For the difficult task of trying to interpret how study findings could have been biased by confounding, researchers require a sound understanding of the underlying principles of confounding, as well as familiarity with methods for identifying potential confounders, such as causal diagrams, and adequate design and analysis strategies for combatting confounding. The aim of this article was to explain in a comprehensible manner the concept of confounding within the counterfactual framework. It may be

helpful as a teaching tool about the basic principles of confounding and the rationale for common design and analytic strategies to handle confounding.

CRediT authorship contribution statement

Martijn J.L. Bours: Conceptualization, Methodology, Visualization, Writing - original draft.

Acknowledgments

The author wishes to thank Prof. Matty Weijenberg and Dr. Eline van Roekel for their helpful feedback and suggestions on earlier drafts of this manuscript.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2020.01.021.

References

[1] Porta M. In: Porta M, editor. A dictionary of epidemiology. 6th ed. New York: Oxford University Press; 2014:377.
[2] Grimes DA, Schulz KF. Bias and causal associations in observational research. Lancet 2002;359:248−52.
[3] Greenland S, Morgenstern H. Confounding in health research. Annu Rev Public Health 2001;22:189−212.
[4] Greenland S, Robins JM. Identifiability, exchangeability and confounding revisited. Epidemiol Perspect Innov 2009;6:4.
[5] Wickramaratne PJ, Holford TR. Confounding in epidemiologic studies: the adequacy of the control group as a measure of confounding. Biometrics 1987;43:751−65.
[6] Greenland S, Robins JM, Pearl P. Confounding and collapsibility in causal inference. Stat Sci 1999;14(1):29−46.
[7] Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. Int J Epidemiol 1986;15:413−9.
[8] Suzuki E, Mitsuhashi T, Tsuda T, Yamamoto E. A typology of four notions of confounding in epidemiology. J Epidemiol 2017;27(2):49−55.
[9] VanderWeele TJ, Shpitser I. On the definition of a confounder. Ann Stat 2013;41(1):196−220.
[10] Miettinen OS, Cook EF. Confounding: essence and detection. Am J Epidemiol 1981;114:593−603.
[11] Miettinen O. Confounding and effect-modification. Am J Epidemiol 1974;100:350−3.
[12] Greenland S, Robins JM. Confounding and misclassification. Am J Epidemiol 1985;122:495−506.
[13] Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology 1999;10:37−48.
[14] Pearl J. An introduction to causal inference. Int J Biostat 2010;6(2). Article 7.
[15] Hernan MA, Clayton D, Keiding N. The Simpson's paradox unraveled. Int J Epidemiol 2011;40:780−5.
[16] Hofler M. Causal inference based on counterfactuals. BMC Med Res Methodol 2005;5:28.
[17] Maldonado G, Greenland S. Estimating causal effects. Int J Epidemiol 2002;31:422−9.
[18] Hernan MA. A definition of causal effect for epidemiological research. J Epidemiol Community Health 2004;58:265−71.
[19] Newman SC. Commonalities in the classical, collapsibility and counterfactual concepts of confounding. J Clin Epidemiol 2004;57:325−9.
[20] Maldonado G. The role of counterfactual theory in causal reasoning. Ann Epidemiol 2016;26:681−2.
[21] Dawid AP. Causal inference without counterfactuals. J Am Stat Assoc 2000;95:407−48.
[22] Vandenbroucke JP, Broadbent A, Pearce N. Causality and causal inference in epidemiology: the need for a pluralistic approach. Int J Epidemiol 2016;45:1776−86.
[23] Hernán MA, Robins JM. Causal inference. Boca Raton: Chapman & Hall/CRC, forthcoming; 2019.
[24] Pearl J. Causality: models, reasoning, and inference. 2th ed. Cambridge University Press; 2009:484.
[25] Lipsky AM, Greenland S. Confounding due to changing background risk in adaptively randomized trials. Clin Trials 2011;8:390−7.
[26] Greenland S. Randomization, statistics, and causal inference. Epidemiology 1990;1:421−9.
[27] Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. Epidemiology 2009;20:488−95.
[28] Rothman KJ, Greenland S, Walker AM. Concepts of interaction. Am J Epidemiol 1980;112:467−70.
[29] de Mutsert R, Jager KJ, Zoccali C, Dekker FW. The effect of joint exposures: examining the presence of interaction. Kidney Int 2009;75(7):677−81.
[30] VanderWeele TJ. On the distinction between interaction and effect modification. Epidemiology 2009;20:863−71.
[31] Greenland S. Interactions in epidemiology: relevance, identification, and estimation. Epidemiology 2009;20:14−7.
[32] VanderWeele TJ, Knol EF. A tutorial on interaction. Epidemiol Methods 2014;3(1):33−72.
[33] Brumback B, Berg A. On effect-measure modification: relationships among changes in the relative risk, odds ratio, and risk difference. Stat Med 2008;27:3453−65.