Confounding variables in epidemiologic studies: Basics and beyond

ARTICLE *in* ARCHIVES OF IRANIAN MEDICINE · AUGUST 2012

Impact Factor: 1.11 · Source: PubMed

1 AUTHOR:

Farin Kamangar
Morgan State University
157 PUBLICATIONS **5,679** CITATIONS

SEE PROFILE

Epidemiology Review

Confounding Variables in Epidemiologic Studies: Basics and Beyond

Farin Kamangar MD•[1,2]

Abstract

This article discusses the importance, definition, and types of confounders in epidemiology. Methods to identify and address confounding are discussed, as well as their strengths and limitations. The article also describes the difference among confounders, mediators, and effect modifiers.

Keywords: Confounder, effect modifier, interaction, mediator, randomization, regression

Cite this article as: Kamangar F. Confounding Variables in Epidemiologic Studies: Basics and Beyond. *Arch Iran Med*. 2012; **15(8):** 508 – 516.

Introduction

In epidemiology, like other fields of science, we look for causes of diseases, by which we mean exposures that change the risk of diseases. For example, by "smoking causes lung cancer", we mean smoking increases the risk of lung cancer; lifetime risk of lung cancer is 17% in male smokers versus 1% in male non-smokers.[1] Once we find that smoking causes lung cancer, people are encouraged not to smoke and public policies are made.

Whereas causation always results in a change in risk, the converse is not necessarily true. Increased risk of a health outcome in the presence of an exposure doesn't necessarily imply a causal relationship between the exposure and outcome. One reason for such non-causal associations is the presence of a third variable called confounder or confounding variable. See the example below.

Example 1: Some epidemiologic studies have found that poor oral health and/or tooth loss is associated with an increased risk of esophageal cancer.[2,3] But does this mean that poor oral health *causes* esophageal cancer? Maybe yes. But maybe there are other factors (e.g., smoking) behind the scene. Smoking causes poor oral health and it also causes esophageal cancer. Therefore, an association between tooth loss (the exposure) and esophageal cancer (the outcome) may be due to smoking (a confounder). ●

In this article, we discuss the following topics:

1) Criteria for confounding;
2) Types confounders;
3) Surrogate confounders;
4) Stratification as a method to understand confounders;
5) Confounders versus other "third" variables (mediators and effect modifiers);
6) Confounding versus selection bias;
7) Confounding by indication;
8) How to identify potential confounders;
9) Methods used to address confounders;
10) Deficiencies of methods used to address confounders;
11) Overadjustment; and
12) How strongly can the confounders distort the associations.

In the final part, summary and conclusions, we tie these 12 topics together and provide a framework for thinking about and handling confounders.

1. Criteria for confounding

A confounder is a variable that distorts the association between two other variables (the exposure and the outcome). Often the exposure is what is being studied as a *potential cause* of the outcome, such as tooth loss in Example 1. Statistical adjustment for the confounder results in a change of relative risk. For a variable to be a confounder, it must have three characteristics: 1) it must be associated with the exposure (causally or not); 2) it must be a cause, or a surrogate of the cause, of the health outcome; 3) it should *not* be in the causal pathway between the potential risk factor and outcome.[4] See Example 2.

Example 2: Research shows that higher parity (mother's number of pregnancies) is associated with higher risk of Down syndrome. For example, on average, the tenth pregnancy is more likely to result in a child with Down syndrome than the first pregnancy. However, we know that this association is not because of parity, but it is because of the age of the mother, as the tenth child is on average born to an older mother than the first child. In fact, the tenth child of a mother who is 26 years old at pregnancy may have a lower risk of Down Syndrome that the first child born to a mother who is 39 years old. In this case age (the confounder) is associated with both the exposure (parity) and the outcome (Down Syndrome) but doesn't come in between them (Figure 1). ●

Figure 1. The association between parity and Down Syndrome is confounded by maternal age. In this figure, double sided arrow denotes association (causal or non-causal), one-sided arrow means causal association, and the dashed arrow denotes a potential causal association that is under investigation.

Authors affiliations: [1] School of Community Health and Policy, Morgan State University, Baltimore, MD, USA, [2]Digestive Disease Research Center, Tehran University of Medical Sciences, Tehran, Iran.
•Corresponding author and reprints: Farin Kamangar MD PhD, Department of Public Health Analysis, School of Community Health and Policy, Morgan State University, Portage Avenue Campus, Room 302, Baltimore, MD 21251
Email: farin.kamangar@morgan.edu
Accepted for publication: 12 June 2012

Example 3: Ginseng is an herb, mainly cultivated in China and Korea, which is used for medicinal purposes. Some people believe that it can strengthen the body and prevent diseases. A cohort study that investigated the association of ginseng with gastric cancer in China found that, contrary to initial expectations, ginseng increased the risk of gastric cancer by 40% (relative risk of 1.40).[5] However, after adjusting for age, the association completely disappeared and ginseng neither increased nor decreased the risk. In this case, age was the confounder; older people were more likely to use ginseng and they were more likely to develop gastric cancer (Figure 2).●

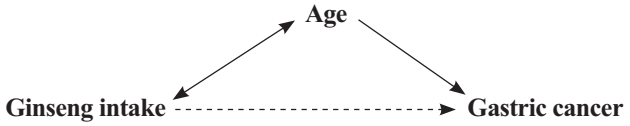

Figure 2. Increased risk of gastric cancer associated with ginseng intake is explained by age.

It is important to evaluate the above-mentioned three requirements before we consider a variable as a confounder. Consider a study of alcohol consumption and breast cancer. Smoking is *not* a confounder in this study. Smoking is related to alcohol consumption but not to the risk of breast cancer.[6] So it does not satisfy all the three requirements.

2. Types of confounders

Confounders may be classified into two categories: qualitative and quantitative. After adjusting for *qualitative* confounders, the association between exposure and outcome completely disappears or even reverses direction, meaning that the quality or nature of the association changes. In examples 2 and 3, the association disappeared after adjusting for age, which was the confounder in both cases. See example 4 for a confounder that reverses the direction of association. Unlike that for qualitative confounders, adjusting for *quantitative* confounders only changes the magnitude of the association but not its nature. See example 5 below for a quantitative confounder.

Example 4: Obesity, sedentary life-style, air pollution, and smoking all make life shorter. So, why is it that people had a much shorter life span 500 years ago, when they were much leaner, were more physically active, were breathing cleaner air, and smoked less? The answer lies in the confounding effect of advances in modern life, such as better hygiene, and development of vaccines and antibiotics. This is an example of confounding that reverses the real association. Had we not adjusted for the advances, comparing now with 500 years ago might have led to a conclusion that was exactly opposite the truth. ●

Example 5: The results of a cohort study in Iran showed that opium consumption was associated with an increased risk of death with a relative risk of 2.26 (126% increased risk).[7] One potential confounder was tobacco use, as tobacco users are more likely to use opium (association with the exposure) and also more likely to die (association with the outcome). Age, sex and other factors may also act as confounders in this association. In fact, after adjusting for smoking, age, sex, and some other potential confounders, this association was less strong (relative risk of 1.86, or 86% increased risk) but it did not disappear. Here, the confounders resulted in only in a change in relative risk. Therefore, they are quantitative confounders.●

Figure 3. The association between opium use and mortality is to some extent confounded by smoking.

Quantitative confounders can further be classified into positive and negative confounders. *Positive confounders* are those that magnify the association beyond its real size – i.e., make the association seem to be bigger than it is. Negative confounders are those that make the association seem to be smaller than it is. Adjustment for positive confounders results in a relative risk that is closer to one, and adjustment for negative confounders results in a relative risk that is further from one. Example 5 describes a positive confounder, as adjustment reduced the relative risk from 2.26 to 1.86. Figure 4 summarizes the classification of confounders.

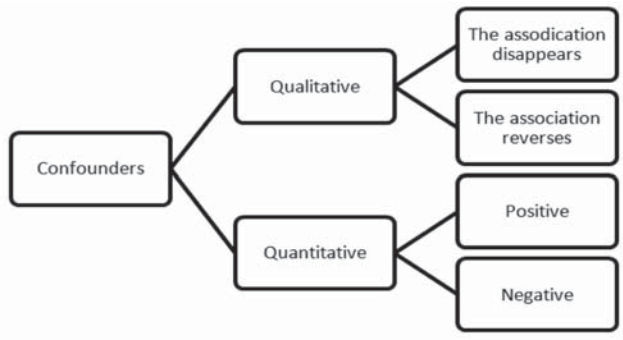

Figure 4. Classification of confounders.

While this terminology (positive versus negative confounders) is sometimes used in epidemiology papers and textbooks,[8] we introduce it here mostly to emphasize that confounders can act in various directions. Learning the concept is more important than the terminology. It is also important to pay attention to the magnitude of change in the relative risk. It is one thing if after adjustment relative risk changes from 6.0 to 5.6, and another thing if relative risk changes from 6.0 to 1.5, although these are both examples of quantitative positive confounders.

3. Surrogate confounders

At times we cannot adjust for the causal confounder itself. In such cases, we may be able to alleviate the problem by adjusting for a variable or a number of variables that together act as a surrogate for the causal confounder. These are surrogate confounders. For example, assume that wealth is a confounder in the relationship between a risk factor called R and an outcome called O. Study participants may have not been asked about their wealth, but they have been asked about their education, their residence zip code, and their profession. A combination of these factors may work as a surrogate for their wealth. In the study of opium and mortality (Example 5),[7] adult height was adjusted for as a surrogate for socioeconomic status during childhood.

4. Stratification as a different method to understand confounding

When a variable acts as a confounder, stratifying the results on the levels of the confounder produces seemingly paradoxical results. The relative risks for each stratum may be different from that seen for the overall association. In the examples below, we illus-

trate the effect of confounding.

Example 6: Assume there is a rare congenital disease called congenia. We conduct a case-control study recruiting mothers of 200 cases of congenia and mothers of 400 control children. A potential risk factor is father's smoking. The table below shows the results for this case-control study.

	Congenia	Controls
Father smoker	140	160
Father not smoker	60	240

From this table, the odds ratio for the association between father being a smoker and congenia is 3.50 (OR = (140 × 240) / (160 × 60) = 3.50).

However, a potential confounder may be mother's smoking, as mother's smoking is related to this disease and men who smoke are more likely to have wives who smoke. When we stratify the results by the two levels of mother's smoking, in neither group do we see an association; i.e., in both groups the odds ratio is 1.00.

In smoking mothers (n = 300)

	Congenia	Controls
Father smoker	135	90
Father not smoker	45	30

OR = (135 × 30) / (90 × 45) = 1.00

In non-smoking mothers (n = 300)

	Congenia	Controls
Father smoker	5	70
Father not smoker	15	210

OR = (5 × 210) / (70 × 15) = 1.00 ●

Whereas we introduce stratification as a method to better understand confounding, it can initially confuse or "confound" the readers. It may take some time and practice before the uninitiated understands this. One may ask "how is it possible that the overall OR is 3.00, but when we stratify the results by whether or not the mother is a smoker, the odds ratio for each group is 1.00? Figure 5 shows this phenomenon.

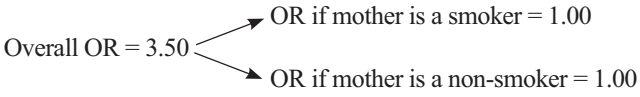

Figure 5. When there is confounding, the overall odds ratio is different from odds ratios in each stratum of the confounder (here, mother's smoking status).

This indeed does confound us! Confounding is part of what is called Simpson's Paradox. The results seem paradoxical, but there is no trick, and it is what happens.

One needs to know that when we stratify the results by the two levels of the confounder, it is unlikely that the two ORs are exactly the same, but as long as they are not statistically significantly different from each other, we usually take a weighted average of the two (using various methods, such as Mantel-Haenszel method). This weighted average is the adjusted odds ratio.

Example 7: The overall OR for the association between X and Y is 3.00. When we stratify the results by the two levels of sex (men and women), the ORs for men and women are 2.20 and 1.90, respectively. Assume that these two numbers are not statistically significantly different from one another (P value = 0.84). If the Mantel-Haenszel weighted average of these two numbers is 2.10, the adjusted OR is 2.10.●

5. Differentiating confounders from other "third" variables (mediators and effect modifiers)

The exposure (the potential risk factor) and the outcome are the two main variables of each association. A confounder is a "third" variable that affects the association of the exposure with outcome. In addition to confounders, there are other "third" variables of interest that play a role in an association. The two most important of such variables are mediators and effect modifiers, and it is important to distinguish confounders from mediators and effect modifiers.

5.1. Mediators versus confounders

One of the characteristics of a confounder, in addition to being associated with the exposure and the outcome, is that it should *not* be in the pathway between the two. If the third variable is in the pathway, it is called a mediator or intermediate factor. See the example below.

Example 8: Poverty is a risk factor for many diseases including myocardial infarction, stroke, diabetes, HIV/AIDS, esophageal cancer, and gastric cancer, to name a few. Let's take the example of poverty and diabetes. Is this association real, or is it confounded by an unhealthy diet? If poverty leads to poor choice of diet or limited access to healthy food, then poverty is a real cause of diabetes and unhealthy diet is a mediator. See Figure 6.●

Figure 6. The association between poverty and diabetes is mediated via limited access to healthy food.

In this case, unhealthy diet is associated with both poverty and diabetes. Also, adjusting for unhealthy diet results in a change in relative risk estimates. But it is not a confounder because it is in the pathway between poverty and diabetes.

A mediator is conceptually different from a confounder. A confounder may result in a non-causal association between the exposure and the outcome, such that an exposure that doesn't cause the outcome is associated with it. Take Example 3. Taking ginseng will not decrease or increase risk of gastric cancer; all of the apparent association is because of age. So ginseng is not a real cause of gastric cancer. However, a mediator simply explains part or all of the reason why the exposure causes that outcome. In this latter case, the exposure really does cause the outcome. For example, poverty causes diabetes. If poverty is eliminated, then unhealthy eating can be reduced, and thus risk of diabetes would be lower. See another example of a mediator below.

Example 9: Having multiple sex partners is a cause of cervical cancer. This causal relationship is mediated through exposure to human papillomavirus (HPV). See Figure 7.●

Figure 7. HPV mediates the causal relationship between having multiple sexual partners and risk of cervical cancer.

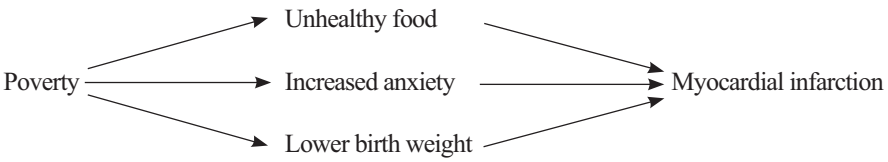

Figure 8. If the effect of poverty on myocardial infarction is mediated via three factors, after fully adjusting for one of these factors, the relative risk shows the effect for the other two factors.

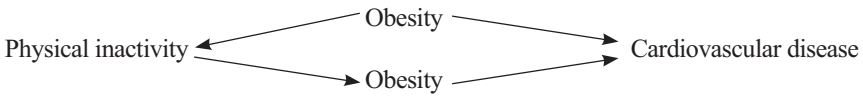

Figure 9. Obesity works both as a confounder and as a mediator in this relationship.

Should we adjust for mediators, as we do for confounders? The answer is that we can, but the meaning of this adjustment is different. Before adjusting for the mediator, we have the total effect of the potential risk factor on the health outcome, whereas after adjusting for the mediator, we have the remaining effect of the risk factor after the partial effect of that mediator is considered. See examples 10 and 11.

Example 10: Assume poverty results in myocardial infarction through three mechanisms: eating more unhealthy food; increasing anxiety; and lower birth weight. Here we have three mediating factors between poverty and myocardial infarction. If we do not adjust for birth weight, then the relative risk (say 2.40) shows the overall effect of poverty on myocardial infarction. However, if we adequately adjust for low birth weight, then adjusted relative risk (say 1.60) will show the effect of poverty on myocardial infarction through the other two mechanisms, i.e., higher anxiety and eating unhealthy food. See Figure 8.●

Example 11: A prospective cohort study of approximately 10,000 civil servants living in London, England, found that those in the lowest socioeconomic position had a 60% increased risk of total mortality (relative risk = 1.60) compared to those in the highest socioeconomic position.[9] After adjusting for several potential mediators, i.e., smoking, alcohol consumption, physical activity, and unhealthy diet, the lowest socioeconomic group were only 14% at higher risk of total mortality (relative risk = 1.14). Therefore, the authors concluded that a substantial fraction of the effect of socioeconomic status on mortality is mediated via these factors.●

While this distinction between confounders and mediators initially seems to be straightforward, in reality, it may be very difficult to determine whether a variable acts as a confounder or as a mediator. Often, it works as both, particularly when there are vicious or virtuous cycles.

Example 12: We want to examine the association between family wealth and risk of overall mortality. Should we adjust for education as a potential confounder? On the one hand, education is associated with both wealth and mortality, and it may not be entirely in the causal pathway. On the other hand, wealthier people are more likely to receive a better education. Therefore, education may be both a confounder and a mediator. Please note that the relationship between education and wealth is one of a virtuous cycle; one leads to the other, and vice versa.●

Example 13: When investigating the association between physical inactivity and cardiovascular outcomes, obesity can act partially as a confounder and partially as a mediator (Figure 9). Obesity due to overeating can make one be less physically active. Also, physical inactivity may lead to obesity and in turn to cardiovascular outcomes. The relationship between obesity and physical inactivity is one of a vicious cycle; one leads to the other, and vice versa. See Figure 9.●

5.2. Effect modifiers versus confounders

Effect modifiers are also third variables that affect the relationship between the exposure and the outcome. A detailed discussion of effect modifiers is beyond the scope of this article. However, we provide a brief treatment here. Effect modifiers are variables that modify the strength of the association between the exposure and the outcome. Stratification is a method to identify effect modifiers. When we stratify the results of the association of a potential risk factor and a health outcome by the two levels of the third variable, if the two relative risks (or the two odds ratios) are statistically significantly different from each other we will conclude that there is effect modification (interaction). See example 14.

Example 14: In the study mentioned in Example 5, the overall adjusted relative risk for the association between opium and overall morality was 1.86. This association was stronger for women (relative risk = 2.43) than for men (relative risk = 1.63); P value < 0.001. This means that while opium increased the risk of death in both men and women, it did so more strongly for women. ●

Please note that to learn about confounding, we compare the adjusted relative risk with the unadjusted relative risk. In contrast, to learn about effect modification (interaction), we compare the relative risks across strata. See examples 15 to 18 and the associated figures.

Example 15: In unadjusted analyses, X is associated with Y with a relative risk of 2.00. When we stratify by sex, the relative risks are 3.00 for women and 1.00 for men (p-value for interaction = 0.001), and the weighted average of these two numbers is 2.00. Here, the effect of X on Y depends on sex; it increases risk in women but not in men. This is a clear example of effect modification by sex. However, the average result is 2.00, after and before adjustment. So, there is not much evidence for confounding by sex. See Figure 10.●

Example 16: In unadjusted analyses, X is associated with Y with a relative risk of 2.00. When we stratify by sex, the relative risks are 1.40 for women and 1.50 for men (p-value for interaction = 0.78), and the weighted average of these two numbers is 1.46. Here, the effect of X on Y does *not* depend on sex; it increases risk in both women and men to almost the same extent. So there is no effect modification by sex. However, the adjusted relative risk

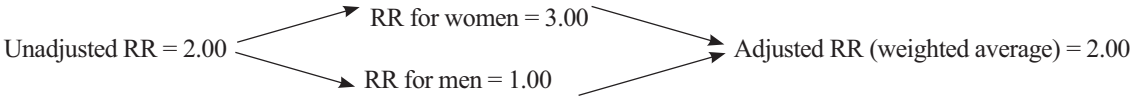

Figure 10. Relative risks across strata of sex differ (effect modification by sex) but the adjusted relative risk is the same as the unadjusted relative risk (no confounding).

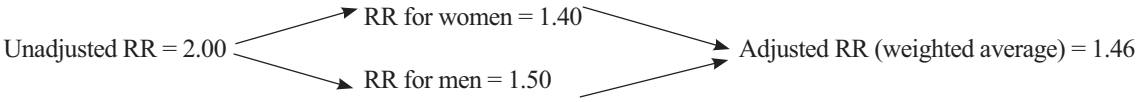

Figure 11. Relative risks are similar across strata of sex (no interaction by sex) but the adjusted relative risk is different from the unadjusted relative risk (confounding).

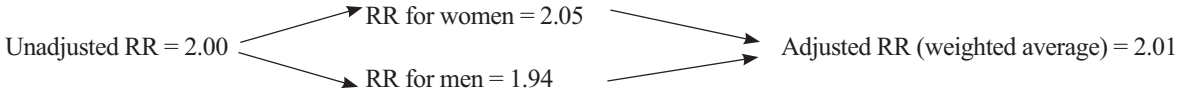

Figure 12. Relative risks are similar across strata of sex (no interaction by sex), and the adjusted relative risk is similar to the unadjusted relative risk (no confounding).

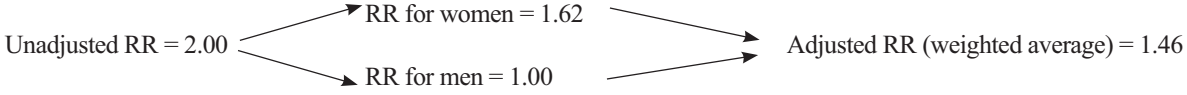

Figure 13. Relative risks depend on sex (interaction by sex) and the adjusted relative risk is different from the unadjusted relative risk (confounding).

(1.46) is different from the unadjusted one (2.00). This is a clear example of confounding by sex. See Figure 11.●

Example 17: In unadjusted analyses, X is associated with Y with a relative risk of 2.00. When we stratify by sex, the relative risks are 2.05 for women and 1.94 for men (*P*-value for interaction = 0.66), and the weighted average of these two numbers is 2.01. Here, the effect of X on Y does *not* depend on sex, so there is no interaction by sex. Also, adjusted and unadjusted relative risks are similar, so there is no confounding by sex. See Figure 12.●

Example 18: In unadjusted analyses, X is associated with Y with a relative risk of 2.00. When we stratify by sex, the relative risks are 1.62 for women and 1.00 for men (p-value for interaction = 0.01), and the weighted average of these two numbers is 1.46. Here, the effect of X on Y depends on sex, so there is interaction by sex. Also, adjusted and unadjusted relative risks are substantially different, so there is confounding by sex. See Figure 13.●

6. Confounding versus selection bias

Some forms of selection bias, such as the difference between the exposed and unexposed in the baseline of a cohort, can be alternatively classified as confounding.

Example 19: In a study of oropharyngeal cancer patients, White patients had a much better survival than Black patients.[10] This difference, however, was shown to be the result of higher prevalence of HPV-induced tumors in White patients with oropharyngeal cancer. On the one hand, this is confounding, because HPV-induced tumors are associated with being White and concomitantly HPV-induced oropharyngeal tumors have better prognosis than other forms of this cancer. On the other hand, this can be considered selection bias, as there is a systematic difference in the type of tumor that Whites and Blacks have. ●

7. Confounding by indication

Confounding by indication is a form of selection bias. This term is used to describe a type of confounding encountered in observational epidemiologic studies of drugs. Since in observational studies the treatment is not dictated randomly – it is rather based on indication for treatment (hence the name confounding by indication) – those who take the drug may be substantially different from those who don't with respect to several characteristics. For example, those who have a severe disease may be more likely to receive the treatment. Therefore, if one finds an association between treatment and higher mortality, it may be due to confounding by indication rather than the adverse effect of the drug. Confounding by indication is most often seen in for drugs that are rarely used but also seen for commonly used drugs such as acetaminophen (Tylenol). Several excellent examples are provided by Signorello et al.[11]

8. How to identify and select potential confounders

One of the major difficulties in epidemiologic studies, particularly in observational studies, is to determine what the potential confounders are, or what to adjust for. There are multiple methods to select confounders[12–16] but these methods can be classified into two broad categories: *a priori* selection methods based on our knowledge of the field, and selection methods based on statistical analysis of the data.

8.1. A priori selection of confounders

In Example 5 (opium and overall mortality), there are a few obvious confounders that one can determine *a priori*. For example, the locals know that in that particular area of Iran older people and men are more likely to use opium, and those are the same people who are at higher risk of death. So adjusting for sex and age is a

must. Similarly, tobacco users are more likely to use opium and more likely to die. So the need to adjust for tobacco use seems to be obvious too. But it is unlikely that we know all the potential confounders or to have collected data on all of them.

Associations vary geographically and change through time. Therefore, confounders may vary by time and population, which may make *a priori* selection of confounders difficult. See the example below.

Example 20: The results of a large prospective cohort study, published in the *New England Journal of Medicine*,[17] showed that coffee drinkers were less educated than non-drinkers and coffee drinking was associated with a number of unhealthy behaviors, including smoking, drinking large amounts of alcohol, physical inactivity, and consuming less fruits and vegetables. Therefore, the association of coffee drinking with total mortality was confounded by these factors. Whereas the unadjusted relative risks showed an increased risk of death associated with coffee drinking, after adjustment the association changed qualitatively and coffee was shown to reduce risk of death. However, this pattern could easily change in 20 years. Educated people are more likely to read the results of new research on health. If more papers like this are published, perhaps in the future we will see that educated people are more likely (rather than less likely) to drink coffee, and coffee drinking may become associated with healthy habits.●

This variation in pattern could pose a challenge for some observational studies. For example, if in some countries tomatoes are heavily exposed to pesticides, while in others they are not, assessing the effect of tomatoes on health could be different across countries because of variation in confounding patterns. If the researchers have the data on all important confounders, in theory, they can adjust for them. However, often this is not the case.

8.2. Statistical selection of confounders

In these methods, the researchers look into the association between a large number of variables (as potential confounders) with both the potential risk factor and the health outcome. If one finds an association with both, then that factor may be a confounder. In the example of opium and mortality, one may explore socioeconomic status, body mass index, ethnicity, marital status, and a number of other variables as potential confounders. A variety of statistical criteria have been used to select the confounders, including change in estimates and statistical significance tests.[12] These methods are deficient in several ways, too. First, in addition to confounders, mediators are associated with both the potential risk factor and the outcome, and it may be difficult to distinguish between the confounders and mediators using statistical methods. Second, testing for a large number of variables may lead to chance findings. Third, there are always a number of unknown or unmeasured potential confounders. Fourth, there is no universal agreement on which of these statistical methods performs best. Fifth, it is difficult to determine the cutpoint for considering a variable a confounder; 5%, 10%, and 20% change in the estimates have all been used, but choosing between these is somewhat arbitrary.

The truth is that neither the *a priori* method nor the statistical methods work perfectly well, and it is impossible to determine all the relevant confounders. However, we can do our best through experience, educated guesses, and trying a combination of variables. Also, as explained above, a combination of several surrogate confounders in the adjustment model may work reasonably well.

9. Methods used to address confounders

Several methods are available to address confounders. Where possible, randomization with large sample sizes is the strongest method to minimize the effect of confounders. As discussed below, randomization with large enough sample sizes can balance the different arms of the study for *both known and unknown* confounders.

In observational studies, where randomization is not possible, a variety of other methods are used to control for confounders. These methods include matching, restriction, stratified analyses, and regression methods. Some of these methods are used during the design of the study (matching and restriction) and some during analysis (stratified analyses and regression methods). This is not an exhaustive list; other methods, such as propensity score matching may also be used.[18] As discussed below, none of the methods used in observational studies are entirely satisfying.

9.1. Randomization

Confounders are major issues in analyzing validity of observational studies, such as case-control or cohort studies. In a cohort study of dietary vitamin E and cancer, for example, individuals who take more vitamin E may be different from those who don't in many ways. In one such cohort study in the United States, those who took more dietary vitamin E were more likely to be female, abstain from smoking, use less alcohol, be physically active, and possess at least an undergraduate college degree.[19]

Compared to observational studies, randomized trials are less subject to confounders, particularly when the sample size is very large. Given that the assignment of the exposure in randomized trials is done at random, it is independent of all characteristics of the participants, such as their age, sex, wealth, etc. Therefore, large randomized trials potentially adjust *for both known and unknown* confounders. For example, in a large randomized trial vitamin E and cancer, the group that received vitamin E was very similar to the group that did not receive it with respect to age, body mass index, cigarette smoking, exercise, alcohol consumption, aspirin use, parental history of cancer, and self-reported history of cancer.[20] This list is only a sample of variables that are similar between the two arms of the study, and one could be highly confident that the two arms are balanced for nearly all other confounders.

However, we should add that small randomized trials are at risk of confounding, as the two groups may be substantially different by chance.[21] Nevertheless, if there are several such small trials, a meta-analysis of them may effectively handle the problem. When it comes to adjustment, for most practical purposes, trials with over 1000 subjects in each arm can be considered large, those with fewer than 100 subjects in each arm may be considered small, and those with 100 to 1000 in each arm are intermediate size.

9.2. Matching

In cohort studies, one can match exposed and unexposed individuals for the potential confounder. For example, if we are assessing the effect of opium on total mortality and sex is a potential confounder, one can match a male opium user to a male opium non-user and a female opium user to a female opium non-user. This way users and non-users will be exactly the same for sex, and thus sex could not confound the association. By extension, one can match for more than one variable, such as by age and sex. For example a 56-year-old male opium user can be matched to a 56-year-old male non-user. However, there is a practical limit to the number of variables that we can match for; it is often difficult

Box 1. The most commonly regression models in epidemiologic studies

The three most commonly used regression models in epidemiology are linear regression, logistic regression, and Cox proportional hazards regression. Linear regression is used mainly when the outcome is continuous (e.g., weight). Logistic regression is mostly used when the outcome is binary (e.g., breast cancer; either the study subject gets breast cancer or not). Extensions of logistic regression, such as polytomous logistic regression and ordinal logistic regression, are also sometimes used. Polytomous logistic models are used for outcomes with three or more categories treated as nominal variables, in which each category is compared to a reference category. For example, risk of three different cancer types (lung, stomach, and esophagus) can be simultaneously compared to a control group. Ordinal logistic regression models are used when the outcome has three or more categories but is treated as an ordinal variable. Cox proportional hazards models are used when the outcome is "time to event", for example time from entering the study to be being diagnosed with lung cancer. In a logistic model, two people who got lung cancer during the follow-up contribute similarly to the outcome, whereas in Cox models, if the first person got lung cancer in one year and the other got lung cancer in 10 years, their contribution is different, as time to event differs. Measuring time to event needs follow-up, therefore Cox regression is often used in prospective studies with follow-up such as cohort studies and randomized clinical trials.

to find a good match based on age, sex, ethnic group, education, wealth, total intake of fruits and vegetables, etc. This makes matching a less useful method than regression (see below).

9.3. Restriction

In this method, we restrict the study population to one level of the potential confounder. For example, if a researcher wants to study the association of opium use and total mortality and is highly concerned about confounding by tobacco use, she can restrict the study population to those who have never used any form of tobacco. This is indeed an extreme form of matching. Restriction can also be done during analysis. The statistician can limit the analysis to a subgroup of all study participants, such as to never-tobacco-users only. Problems with restriction are even more severe than matching. It is difficult to restrict the study population to a group based on several variables (age, sex, ethnic group, education, etc.), as sample size becomes small and generalizability of the results will be limited. Again, using regression methods is favorable.

9.4. Stratified analysis

In this method, the statistician stratifies the analysis on different levels of the potential confounder to examine if there is evidence for confounding. For example, if sex is a potential confounder, the statistician can analyze the results separately by male and female. Like the two previous methods, stratification has a practical limit for the number of variables chosen to stratify. For example, if we choose sex (two levels: male and female) and race (four levels: White, Black, Asian, Other), the data need to be stratified into $4 \times 2 = 8$ strata. If we additionally stratify based on 10 categories of age, data will have $4 \times 2 \times 10 = 80$ strata. Again, this favors using regression methods. However, as explained above, stratification is important in understanding confounding and to distinguish it from effect modification.

9.5. Regression

Multipredictor regression models adjust for confounders by modeling the exposure and potential confounders in relation to the outcome. These regression models estimate the effect of the exposure while keep the levels of the confounder constant. For example, if the potential confounder is sex, regression models act as if they estimate the effect of the exposure for men and women separately and take a weighted average of the results. This is intuitively similar to stratification but offers several advantages over stratification. See below. Depending on the type of outcome, several regression models can be used. See Box 1.

9.6. Choice of method of adjusting for confounders

Where possible, randomization with large sample sizes is the most effective method for dealing with confounders, as it balances the different arms of the study for both known and unknown confounders. However, due to ethical and logistic problems, randomization is often not feasible. For reasons mentioned above, the most commonly used method to handle confounders in observational studies is using multi-predictor regression methods. These methods are able to control for several confounders at a time and put relatively little restriction on the study design and participant recruitment.

In theory, matching can be used in small randomized studies and cohort studies to control for confounding, but this is rarely done in such studies. Matching is commonly done in case-control studies. However, as Rothman and coauthors have shown, in case-control studies matching may not only be ineffective in dealing with confounders, it may actually cause a new form of confounding.[4] (We understand that this is counter-intuitive. See the reference[4] for further information.) When the matching factor is strongly associated with the exposure but not with the outcome (hence not a confounder), matching may cause confounding. Thus, it is highly recommended that the matching factor be adjusted for in case-control analyses.[4]

Stratified analyses may be illustrative at times and they can help researchers distinguish between confounding and effect modification (discussed above).

Restriction in design is rarely used to control for confounders. When study participants are restricted to a group, it is often for reasons other than confounding, such as for efficiency or ethical reasons. Restricting the analysis to a certain subgroup, such as non-smokers, which is in fact a form of stratified analysis, is often done. In a study of oral health and esophageal cancer, for example, the researchers restricted the analysis to never smokers and still found an association.[3] This was done to avoid any potential for residual confounding from smoking. See below for further information on residual confounding.

10. Deficiencies of methods to address confounders

As mentioned earlier, other than randomization, other methods are not capable of adequately handling confounders. One major problem with all of these methods is that there may be several unknown or unmeasured confounders in each study. Another problem is that even when confounders are known or measured, they may have been measured poorly, leading to inadequate adjustment for them. This latter case is also called residual confounding. See examples 21 and 22.

Example 21: Income is a potential confounder in a study and thus

participants are asked to report their income, but they fail to do so accurately. In this situation, errors in the recorded values of income lead to imperfect adjustment for it, and hence residual confounding.●

Example 22: Diet is usually measured using food frequency questionnaires in epidemiologic studies. However, answers to these questionnaires are subject to substantial misclassification and measurement errors.[22] People barely remember how many tomatoes they have eaten each week during the past year. Even if they do, their diet might have changed over the years. So, the responses do not fully reflect the life-long exposure to that dietary factor. Therefore, although many epidemiologic studies claim that they have adjusted for dietary factors, that adjustment is often quite inadequate. ●

At times potential confounders are measured in broad categories, for example, education may be measured as illiterate, elementary, middle or high school, and higher education. This may also lead to residual confounding. Finally, poor modeling in regression analyses may result in inadequate adjustment and residual confounding. See the Figure 14.

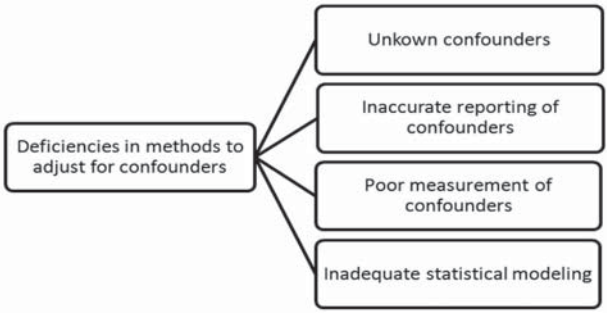

Figure 14. Deficiencies in methods used to adjust for confounders may results in inadequate adjustment.

11. Overadjustment

While we adjust for potential confounders, it may be counterproductive to adjust for too many variables. First, if we adjust for mediators, rather than confounders, the adjusted result will not be a correct estimate of the entire effect of the exposure on the outcome. Second, if one predictor variable in the regression model is highly correlated with another predictor variable, or a combination of variables, we may face a problem called collinearity. When collinearity exists, standard errors of estimates will be very large and estimating the effect of collinear variables will be very imprecise. To maintain precision, sometimes the statistical software drops one of the collinear variables. Extreme cases of collinearity are unusual but they do happen. For example, hemoglobin and hematocrit are highly correlated and therefore collinear, and putting both in the regression model as predictors may cause a problem. Likewise, assume a researcher wants to learn about predictors of infant mortality. If she puts weight, height, head circumference, and abdominal circumference at birth plus gestational age, there will likely be collinearity, as weight at birth should be strongly correlated with a combination of the other four variables. Third, adding a large number of variables (that are not real confounders) to the model for adjustment may slightly reduce power or make the model unstable, particularly if some of the variables are categorical with sparse numbers in some of their categories, and when the ratio of the number of variables included to the sample size is large.

12. How strongly can confounders distort the association?

Confounders should be strongly associated with both the exposure and the outcome to have a material effect on the relative risk estimates. For example, if the risk factor increases the risk of the outcome by 2-fold (relative risk = 2.00), the confounder should be associated with a 5-fold increased risk of both risk factor and outcome to completely negate the association after adjustment. By comparison, in this same example if the confounder is associated with the risk factor with a relative risk of 1.2 and with the outcome with a relative risk of 1.4, it is unlikely to have a material effect on the relative risk; it may just change the risk from 2.00 to say 1.95.

In Example 6, after adjustment for mother's smoking, the unadjusted OR of 3.50 for father's smoking and congenia changed to 1.00. This is a substantial change in OR. However, please note that the association between mother's smoking and congenia (OR = 21.00) and mother's smoking with father's smoking (OR = 9.00) were very strong, both much higher than 3.50, otherwise we wouldn't have seen such a substantial change. These are shown in the tables below.

	Congenia	Controls
Mother smoker	180	120
Mother not smoker	20	280

$$OR = (180 \times 280) / (20 \times 120) = 21.00$$

	Mother smoker	Mother not smoker
Father smoker	225	75
Father not smoker	75	225

$$OR = (225 \times 225) / (75 \times 75) = 9.00$$

Siemiatycki and colleagues conducted an empiric investigation of occupational exposures and various cancers to determine the effect of inclusion and exclusion of three potentially important confounders, i.e., smoking, ethnicity, and socioeconomic status. Of the 75 OR's examined in this study, only eight OR estimates were distorted by more than 20%, of which seven involved lung cancer, a disease very strongly associated with smoking. Therefore, these investigators concluded that *"relative risks between lung cancer and occupation in excess of 1.4 are unlikely to be artifacts due to uncontrolled confounding. For bladder cancer and stomach cancer, the corresponding cut point may be as low as 1.2"*.[23] While not everyone may agree with this latter conclusion, the findings of this study corroborate the fact that not-so-strong confounders are unlikely to change the results substantially.

The level of residual confounding depends on the number of unmeasured or incorrectly measured confounders, their strength of association with exposure and outcome, the prevalence of confounders, and the correlation among confounders. A simulation study showed that under reasonable circumstances, for example when two independent confounders each increase the risk of the outcome by 2-fold, unadjusted odds ratios of approximately 2.00 can be generated (unmeasured or unadjusted confounders).[24] This study showed that under similar circumstances odds ratios of 1.50 can be generated even after adjustment for the two confounders if the confounder is poorly measured (residual confounding). However, the results of this study implied that odds ratios of 2.50 or higher are unlikely to be due to confounding alone.

Change in the magnitude of associations due to confounding is a very important point in discussions of epidemiologic findings. When the initial studies of smoking and lung cancer were published in the 1950s, Ronald Fisher, a world-renowned geneticist and the most prominent statistician of his time, argued that these associations may be confounded by genes;[25] some genes could

cause you to smoke and those same genes could cause lung cancer. Today we know that Fisher, who was a life-time smoker, was wrong in this case. The relative risk for the association is between smoking and lung cancer can be as high as 30 for long-time smokers. If Fisher were to be right, then the relative risks for the association between genes and lung cancer (and genes and smoking) should be substantially higher than 30. This is clearly not the case, as the relative risk for most common polymorphisms associated with lung cancer is very low, around 1.25.[26]

Summary and conclusions

Confounding factors pose a major problem in identifying the real causes of diseases. Confounders may erroneously increase or decrease the magnitude of an association, or even invert the direction of the association. The steps outlined below may help in thinking about and addressing confounders in epidemiologic studies.

Step 1. Do we need to be concerned about confounders? If the study is a large randomized trial, confounding is not a major concern. If it is not, then we need to identify, select, and adjust for potential confounders.

Step 2. How do we find the potential confounders? Potential confounders are found based on *a priori* reasons or statistical reasons. Not all "third variables" are confounders. Assess whether a variable meets the three criteria for confounding. If it does, then also consider how strongly it is associated with the exposure and the outcome and how much it changes the relative risk. Confounders should be strongly associated with both the exposure and the outcome to have a material effect on the results. Otherwise, they may not be of major concern. When confounders cannot be measured or are poorly measured, we need to find surrogate confounders and adjust for them.

Step 3. How do we adjust for confounders? Regression analysis is the most common method to adjust for confounders in observational studies. We should use the most appropriate statistical regression method (e.g., linear regression, logistic regression, Poisson regression, or Cox proportional hazards regression) to adjust for confounders. This choice often depends on the type of the outcome.

Step 4. How do we interpret the results after adjustment? Even after adjusting for confounders, we need to keep in mind that the results may not have been fully adjusted for due to unmeasured or poorly measured confounders (residual confounding). However, often adjusting for surrogate confounders may alleviate the problem. We should be reasonably cautious but not over-critical. Whether or not confounders have been adequately dealt with is a matter of opinion. But experience helps in making educated guesses about the presence and magnitude of residual confounding.

Acknowledgements

The author thanks Dr. Ashkan Emadi (School of Medicine, University of Maryland, Baltimore, MD), Dr. Mahsa Mohebtash (Union Memorial Hospital, Baltimore, MD), Dr. Payam Sheikhattari (School of Community Health and Policy, Morgan State University, Baltimore, MD), and Ms. Gillian Silver (School of Community Health and Policy, Morgan State University) for reading the paper thoroughly and providing constructive comments.

References

1. Villeneuve PJ, Mao Y. Lifetime probability of developing lung cancer, by smoking status, Canada. *Can J Public Health*. 1994; **85:** 385 – 388.
2. Abnet CC, Qiao YL, Mark SD, Dong ZW, Taylor PR, Dawsey SM. Prospective study of tooth loss and incident esophageal and gastric cancers in China. *Cancer Causes Control*. 2001; **12:** 847 – 854.
3. Abnet CC, Kamangar F, Islami F, Nasrollahzadeh D, Brennan PJ, Aghcheli K, et al. Tooth loss and lack of regular oral hygiene are associated with higher risk of esophageal squamous cell carcinoma in a case-control study conducted in Golestan Province, Iran. *Cancer Epidemiol Biomarkers Prev*. 2008; **17:** 3062 – 3068.
4. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. Philadelphia, PA: Lippincott Williams and Wilkins; 2008.
5. Kamangar F, Gao YT, Shu XO, Kahkeshani K, Ji BT, Yang G, et al. Ginseng intake and gastric cancer risk in the Shanghai Women's Health Study cohort. *Cancer Epidemiol Biomarkers Prev*. 2007; **16:** 629 – 630.
6. Hamajima N, Hirose K, Tajima K, Rohan T, Calle EE, Heath CW Jr., et al. Alcohol, tobacco and breast cancer--collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. *Br J Cancer*. 2002; **87:** 1234 – 1245.
7. Khademi H, Malekzadeh R, Pourshams A, Jafari E, Salahi R, Semnani S, et al. Opium use and mortality in Golestan Cohort Study: prospective cohort study of 50 000 adults in Iran. *BMJ*. 2012; **344:** e2502. doi: 10.1136/bmj.e2502.
8. Szklo M, Nieto FJ. *Epidemiology: Beyond the Basics*. 2nd ed. Sudbury, MA: Jones and Bartlett Publishers; 2007.
9. Stringhini S, Sabia S, Shipley M, Brunner E, Nabi H, Kivimaki M, et al. Association of socioeconomic position with health behaviors and mortality. *JAMA*. 2010; **303:** 1159 – 1166.
10. Settle K, Posner MR, Schumaker LM, Tan M, Suntharalingam M, Goloubeva O, et al. Racial survival disparity in head and neck cancer results from low prevalence of human papillomavirus infection in black oropharyngeal cancer patients. *Cancer Prev Res (Phila)*. 2009; **2:** 776 – 781.
11. Signorello LB, McLaughlin JK, Lipworth L, Friis S, Sorensen HT, Blot WJ. Confounding by indication in epidemiologic studies of commonly used analgesics. *Am J Ther*. 2002; **9:** 199 – 205.
12. Maldonado G, Greenland S. Simulation study of confounder-selection strategies. *Am J Epidemiol*. 1993; **138:** 923 – 936.
13. Mickey RM, Greenland S. The impact of confounder selection criteria on effect estimation. *Am J Epidemiol*. 1989; **129:** 125 – 137.
14. Greenland S. Invited commentary: variable selection versus shrinkage in the control of multiple confounders. *Am J Epidemiol*. 2008; **167:** 523 – 529.
15. Miettinen OS, Cook EF. Confounding: essence and detection. *Am J Epidemiol*. 1981; **114:** 593 – 603.
16. Greenland S, Neutra R. Control of confounding in the assessment of medical technology. *Int J Epidemiol*. 1980; **9:** 361 – 367.
17. Freedman ND, Park Y, Abnet CC, Hollenbeck AR, Sinha R. Association of coffee drinking with total and cause-specific mortality. *N Engl J Med*. 2012; **366:** 1891 – 1904.
18. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med*. 1997; **127:** 757 – 763.
19. Carman S, Kamangar F, Freedman ND, Wright ME, Dawsey SM, Dixon LB, et al. Vitamin E intake and risk of esophageal and gastric cancers in the NIH-AARP Diet and Health Study. *Int J Cancer*. 2009; **125:** 165 – 170.
20. Gaziano JM, Glynn RJ, Christen WG, Kurth T, Belanger C, MacFadyen J, et al. Vitamins E and C in the prevention of prostate and total cancer in men: the Physicians' Health Study II randomized controlled trial. *JAMA*. 2009; **301:** 52 – 62.
21. Rothman KJ. Epidemiologic methods in clinical trials. *Cancer*. 1977; **39:** 1771 – 1775.
22. Schatzkin A, Subar AF, Moore S, Park Y, Potischman N, Thompson FE, et al. Observational epidemiologic studies of nutrition and cancer: the next generation (with better observation). *Cancer Epidemiol Biomarkers Prev*. 2009; **18:** 1026 – 1032.
23. Siemiatycki J, Wacholder S, Dewar R, Cardis E, Greenwood C, Richardson L. Degree of confounding bias related to smoking, ethnic group, and socioeconomic status in estimates of the associations between occupation and cancer. *J Occup Med*. 1988; **30:** 617 – 625.
24. Fewell Z, Davey SG, Sterne JA. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *Am J Epidemiol*. 2007; **166:** 646 – 655.
25. Fisher RA. Lung cancer and cigarettes. *Nature*. 1958; **182:** 108.
26. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am J Hum Genet*. 2009; **85:** 679 – 691.