AOGS
Acta Obstetricia et Gynecologica Scandinavica

AOGS REVIEW

An overview of confounding. Part 1: the concept and how to address it

PENELOPE P. HOWARDS

Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA

Key words
Bias, causality, confounding factors (epidemiology), data analysis, epidemiologic methods, epidemiologic research design

Correspondence
Penelope Howards, Department of Epidemiology, 1518 Clifton Rd. NE, CNR 3029, Atlanta, GA 30306, USA.
E-mail: penelope.howards@emory.edu

Conflict of interest
The author has stated explicitly that she has no conflicts of interest with this article.

Please cite this article as: Howards PP. An overview of confounding. Part 1: the concept and how to address it. Acta Obstet Gynecol Scand. 2018; 97:394–399.

Received: 2 October 2017
Accepted: 2 January 2018

DOI: 10.1111/aogs.13295

Abstract

Confounding is an important source of bias, but it is often misunderstood. We consider how confounding occurs and how to address confounding using examples. Study results are confounded when the effect of the exposure on the outcome, mixes with the effects of other risk and protective factors for the outcome. This problem arises when these factors are present to different degrees among the exposed and unexposed study participants, but not all differences between the groups result in confounding. Thinking about an ideal study where all of the population of interest is exposed in one universe and is unexposed in a parallel universe helps to distinguish confounders from other differences. In an actual study, an observed unexposed population is chosen to stand in for the unobserved parallel universe. Differences between this substitute population and the parallel universe result in confounding. Confounding by identified factors can be addressed analytically and through study design, but only randomization has the potential to address confounding by unmeasured factors. Nevertheless, a given randomized study may still be confounded. Confounded study results can lead to incorrect conclusions about the effect of the exposure of interest on the outcome.

Abbreviation: IPW, inverse probability weighting.

Introduction

Confounding is one of the three types of bias that may affect epidemiologic studies; the others being selection bias and information bias (misclassification and measurement error). Although confounding is commonly referenced in research, it is often misunderstood. In this paper, we revisit confounding in order to clarify when it occurs and to discuss how to address it. In a companion paper published in this journal (1), we describe how to identify when confounding is a problem, and we introduce some special types of confounding.

Confounding has been described as a confusion of effects (2). In other words, the effect of the exposure of interest [for example, dietary magnesium supplementation (3)] on the outcome (for example, preterm birth) is confused with the effect of another risk or protective factor for the outcome (for example, maternal age). To draw appropriate conclusions about the effect of the exposure on the outcome, we must separate its causal effect from that of the other factors that affect the outcome. For example, suppose taking magnesium supplements decreases the risk of preterm birth. Further, suppose older mothers (for example, ≥35 years old vs.<35 years old) are more likely both to take magnesium supplements and to have a preterm birth. If maternal age is ignored, the

Key message

Confounding occurs when the exposure effect mixes with effects of other risk factors for the outcome differentially across the exposed and unexposed. Confounding by known factors can be addressed through study design or analytically, but confounding by unmeasured factors may remain.

protective effect of taking magnesium supplements on preterm birth will appear weaker than it actually is because magnesium use is more common among women who are also at higher risk of preterm birth. Numeric illustrations that parallel this paper are provided in Supporting Information Appendices S1 and S2.

Ideal but impossible study design

To definitively know whether a factor causes or prevents an outcome in a given individual, we would need to travel to a parallel universe. For example, if we want to know whether taking magnesium supplements prevents preterm birth, parallel universes would allow us to compare what happens to a given women in one universe where she takes magnesium and in another universe where she does not take magnesium. Everything is the same in these universes except whether the woman takes magnesium or not and the consequences of that action. If the birth outcome in the two universes is different (in one universe she has a preterm birth and in the other she does not), then the difference would be due to magnesium. If the outcome is the same in both universes (she has a preterm birth or has a term birth in both universes) then magnesium supplementation has no effect. In this ideal study, there would be no confounding because the factors, other than magnesium, that affect the outcome do not differ in the two universes. For example, her age would be the same in both universes so the effect of age (the effect of being older vs. younger) is not mixed with the effect of magnesium supplementation.

This ideal, but impossible, study design can be extended to the population level. Here, we compare a cohort of women who all took magnesium supplements with the same cohort of women in a parallel universe where everything is the same except that none of the women took magnesium. If the risk of preterm birth differs in the two universes, the difference would be attributable to magnesium supplementation. The women in this ideal study could have a broad range of ages, but maternal age would not confound the study results because each individual woman would be the same age in each universe. Thus, any effect of age on her pregnancy would be the same in each universe. Only her magnesium supplementation status would change, so only that could change her pregnancy outcome. We could calculate the true, causal effect of magnesium supplementation on preterm birth by comparing the risk of preterm birth from the exposed universe with the risk from the unexposed universe.

Unfortunately, observing parallel universes is not a viable study design. Instead, we attempt to mimic this design by picking a substitute population to represent what we cannot observe (4). Suppose our target population, the population about which we want to draw conclusions, is the exposed group in our universe (the women who use magnesium supplements). We observe the pregnancy outcomes of these women so we can calculate the risk of preterm birth in the exposed universe. However, we do not know what would have happened if these same women had not taken magnesium supplements (the risk of preterm birth in the unobserved parallel universe where they were unexposed). Instead, we must identify a different group of women in our observed universe who represent this unobserved experience. Unfortunately, the unexposed women in our actual study may have different characteristics from the exposed women, which means they might have a different risk of the outcome from the exposed women in the parallel universe where they were unexposed. For example, the women in our study who did not use supplements might be younger on average than our target population (the women who used supplements). As a result, the observed risk of preterm birth among the unexposed women in our study would be different from the unobserved risk of preterm birth in the parallel universe where our target population was unexposed because younger women are at a lower risk of preterm birth. Thus, in our actual study, the age difference between the exposed and unexposed women causes the effect of magnesium supplementation to be mixed with the effect of age, unlike the ideal study, where the women are the same age in both universes. This mixing of effects is confounding. Of note, confounding can make an exposure appear to have a stronger, weaker or opposite effect on the outcome than it truly does. Confounding in any direction is important to the degree that it results in erroneous conclusions about the effect of the exposure on the outcome.

This ideal study design, often referred to as potential outcomes, or counterfactual theory, is discussed more formally elsewhere (2,5,6). Although we cannot perform the ideal epidemiologic study, we can address confounding analytically and through study design. However, thinking through the ideal but impossible study design first, helps us make better choices about how to address confounding.

Addressing confounding through study design

Cross-over study design

The (exposure) cross-over study design attempts to mimic the ideal study by observing the outcome for each study participant when exposed and when unexposed. Like the ideal study design, fixed risk factors for the outcome, such as genetic factors, do not change over time and therefore are the same when the woman is exposed

and when she is unexposed. Thus, fixed factors do not confound the results. However, unlike the ideal study design, confounding can occur if there are risk factors that change over time or if the first exposure or first outcome affects the second outcome. For example, in a cross-over study of magnesium supplementation and preterm birth, if all study participants took magnesium during their first pregnancy and did not use magnesium during a second pregnancy, genetic risk factors for preterm birth would not introduce confounding because the women's genes would not have changed. However, there could be confounding by maternal age because the women's ages would be different for the two pregnancies. Thus, even though each woman is exposed and unexposed, the second pregnancy outcome may not be the same outcome she would have had for her first pregnancy in a parallel universe where she did not take supplements. Additional examples, including a more realistic example of a cross-over study, are described in Supporting Information Appendix S3.

Twins study design

A second study design that attempts to mimic the ideal study would be a study of monozygotic twins with discordant exposure status, such as a study of pregnant monozygotic twins where one twin used magnesium supplements and the other did not. Although the exposed and unexposed population would not include the same people, the two populations would include individuals with the same genetic makeup and (usually) the same childhood exposures [for instance, childhood socio-economic status (SES)], which would prevent confounding by these factors. However, confounding could occur because of other factors that differed during childhood or adulthood. For example, if the twins became pregnant at different ages, the pregnancy outcome for the unexposed twin might not represent the pregnancy outcome the exposed twin would have had, had she not taken magnesium (in a parallel universe). However, the unexposed twin might be better at representing this experience than an unrelated pregnant woman, who would have different childhood experiences and different genes.

Randomized study design

Unlike the ideal study design, in studies where exposure is randomized, the exposed and unexposed study participants are different people. However, the expectation is that on average, across repeated trials (or in an infinitely large study), the distribution of risk factors for the outcome will be balanced between the exposed group and the unexposed group at baseline. In other words, if we randomized women to take or not take magnesium supplements, the age distribution among those randomized to magnesium and those randomized to no magnesium would be similar on average, which would limit confounding by maternal age. Theoretically, the women randomized not to receive magnesium would be expected to do a good job estimating the risk of preterm birth the women randomized to take magnesium would have had, if they had not taken magnesium (in a parallel universe). A useful feature of randomization is that, on average, randomization should address confounding by both known and unknown factors. However, in a given study, confounding may still be present. For example, by chance, more older women might be randomized to take magnesium supplements, which would result in confounding by maternal age. As the study size increases, the likelihood of strong baseline confounding occurring by chance decreases. However, post-randomization sources of bias can still occur (7).

Observational study designs

In observational studies, the exposed and unexposed groups are different people and exposure status is determined by the participants themselves or their circumstances. Because exposure status is often correlated with other factors (for example, magnesium supplementation and age), the potential for confounding is increased, and the choice of an appropriate comparison group is critical. Even in observational studies, potential confounding can be addressed by picking study designs that increase the probability that the unexposed group does a good job of representing the experience the exposed group would have had if unexposed. However, these approaches can only address confounding by known risk factors for the outcome of interest.

Restriction

Restriction can prevent confounding by known risk factors for the outcome. Both the exposed and unexposed population are restricted to one level of a factor. Because the factor is present for both the exposed and the unexposed study participants or absent for both groups, the effect of the factor on the outcome is no longer confused with the effect of the exposure of interest. For simplicity, assume it is appropriate to dichotomize women into younger and older women. If a study is restricted to older mothers, both women taking magnesium supplements and women not taking supplements would be older. Therefore, there would be no confounding by maternal age because, like the ideal study design, age would be the same among the exposed and unexposed. Thus, restriction improves the chances that

the risk of preterm birth among study participants who did not take supplements estimates the risk the study participants who took supplements would have had if they had not taken supplements (in a parallel universe). However, this approach only addresses confounding by the restriction factor, so the results may be confounded by other factors. For example, women with a lower income might be less able to afford magnesium supplements and might be more likely to have a preterm birth, so there still could be confounding by income.

Matching

An alternative to restriction is matching. Matching in a cohort study may decrease confounding by improving the probability that the unexposed population represents the experience that the exposed population would have had if unexposed. Suppose women taking magnesium supplements are more likely to be older (≥35 years old vs. <35 years old) than women not taking supplements (for example, 20% older among those taking supplements vs. 10% older among those not taking supplements). To address this, we could match the proportion of older women who did not take magnesium in our study to the proportion of older women who did take magnesium by matching individual women taking magnesium supplements to women in the same age category who are not taking magnesium supplements, or by oversampling older women not taking supplements without individual matching. Using either method to ensure that 20% of women are older among both those using and not using magnesium supplement would decrease confounding by maternal age. Thus, matching increases the probability that the risk of preterm birth for the women who did not take magnesium in our study estimates the risk of preterm birth that the women who did take magnesium in our study would have had if they had not taken magnesium (in the parallel universe). However, as with restriction, this only addresses confounding by the matching factor or factors, not confounding by other unmatched factors, such as income.

Matching in a cohort study addresses confounding by forcing the distribution of the matching factors to be the same among the exposed and unexposed group. However, matching in a case-control study forces the distribution of the matching factor to be the same among those with and without the outcome (preterm and term births), which does not parallel the ideal, confounding-free study design with an exposed and an unexposed universe. Because a confounder is associated with the exposure, matching cases and controls on a confounder causes the distribution of the exposure among the controls to be more similar to the distribution of the exposure among the cases than it would be in the source population. However, the controls are supposed to represent the distribution of the exposure in the source population. Thus, matching on a confounder introduces a form of selection bias, which tends to bias the unadjusted odds ratio towards the null regardless of the true direction of the association, because it causes cases and controls to have a more similar distribution of the exposure than they do in the source population (2). This selection bias and the original confounding can be addressed by adjusting for the matching factor analytically (see below).

Thus, the motivation for matching in a case-control study is to improve efficiency rather than to address confounding. Generally, confounding in a case-control study can be addressed analytically by many of the methods described below. However, if a confounder is common among the cases but rare among the controls (or vice versa), and resources limit the number of controls that can be recruited, then matching avoids having strata of the confounder where the number of cases and controls are extremely imbalanced, which improves statistical efficiency (decreases the estimated variance).

Addressing confounding through analysis

Although a thoughtful study design can decrease confounding, most studies also require confounding to be addressed analytically. We discuss a few of the most commonly used methods and how they relate to the ideal study design below. We start by considering examples of how to adjust for a single confounder, but address adjusting for multiple confounders in the section on modeling.

Stratification

Stratification is analogous to restriction. For this approach, separate estimates of the effect of the exposure on the outcome are calculated for each strata of the covariate. For example, a separate analysis would be done to estimate the effect of magnesium supplementation on preterm birth for older and for younger mothers. The effect of magnesium supplementation would not be confounded by maternal age in each stratum because maternal age group would not vary within a given stratum. Theoretically, this approximates an ideal study among older mothers and one among younger mothers, assuming dichotomizing maternal age is appropriate and that there are no other confounders or other sources of bias.

Standardization

We can also approximate the ideal study design through standardization by analytically forcing the distribution of

the confounder to be the same in the unexposed group as it is in the exposed group. Standardization is analogous to matching, except that it occurs after the data are collected. By default, the risk of preterm birth in the total population is a weighted average of the risk of preterm birth among the older women and the risk among the younger women. The weights are typically different among the exposed (20% are older) and the unexposed (10% are older). Standardization changes the weights for the exposed and unexposed populations so they are same (Appendix S1). When the target population is the exposed group, the weights from the exposed population are used (20% older and 80% younger). Thus, standardization weights the risk of having a preterm birth among older women who did not use magnesium heavier (up from 10 to 20%) to match the weight among supplement users and weights the risk of preterm birth among younger women who did not use magnesium down (from 90% to 80%). In other words, this calculation determines what the risk in the unexposed group would have been if it had had the same age distribution as the exposed group; we hope that this represents the experience the exposed group would have had in the parallel universe where they were unexposed. The ratio of the standardized risks for the exposed and unexposed groups estimates the standardized effect estimate, which is not confounded by age. This approach to standardization is a form of direct standardization, and it approximates the ideal study design where the target population has the same age distribution in both universes.

Indirect standardization is another calculation that can be used to address confounding, but it is not typically used in studies where an exposed and unexposed population have been recruited. The standardized mortality ratio is a common example of indirect standardization.

Pooling

Pooling also forces the distribution of the confounder (for example, maternal age) to be the same among the exposed and unexposed group. There are various methods of pooling, including inverse variance weighting and the Mantel–Haenszel odds ratio. Although similar to standardization, there is a subtle difference between these methods. Standardization forces the distribution of the confounder in the unexposed group to be the same as the distribution in the exposed group, which improves the chances that the unexposed group represents the experience of the exposed if they had been unexposed. In contrast, the weights in pooling are proportional to the amount of information available in each stratum of the confounder. For example, if there are more younger women than older women overall, then the stratum with younger women receives more weight even if being an older mother is more common in the target population (i.e. the exposed). Therefore, pooling is less directly comparable to the ideal study design with the exposed group as the target population, although it still makes the exposed and unexposed group more similar than they would be without adjustment.

Inverse probability weighting

Inverse probability weighting (IPW), described in greater detail elsewhere (8), is directly analogous to the ideal study if the target population were the total population rather than the exposed group. Essentially, this method creates a pseudo-population that represents what would happen if everyone in the study were exposed and everyone unexposed. Because each study participant is included twice, the distribution of the confounder is the same for the exposed and the unexposed groups. The pseudo-population is created by applying weights to the observed study population, as described in Appendix S1. Essentially, the weights assign outcomes to the hypothetical members of the pseudo-population based on the assumption, for example, that the unexposed older women represent the experience that all the older women would have if they were unexposed and that the exposed older women represent the experience that all older women would have if they were exposed. Parallel assumptions are made for the younger women. Like the ideal study design, confounding by factors used in calculating the weights is removed. However, confounding by unmeasured factors may remain.

Modeling

The analytic methods to address confounding described thus far are easiest to apply when there are a limited number of categorical, confounding variables (although IPW can be performed using a more flexible modeling approach). Addressing confounding through modeling can accommodate multiple confounders at one time and is not restricted to categorical variables. Including a confounder in a traditional regression model is analogous to pooling. The effect estimate for the exposure of interest is interpreted as the estimated effect when the confounder is held constant (for example, when maternal age is held constant as older or younger). If no interaction term is included in the model, the estimated effect of magnesium supplementation on preterm birth is assumed to be the same for older and younger mothers.

Alternatively, IPWs can be calculated using models that include multiple confounders (for example, maternal age and income) to predict exposure status (magnesium supplement use). These weights are then applied in a model

to estimate the effect of the exposure (magnesium supplement use) on the outcome (preterm birth), as described in more detail elsewhere (9).

Summary

Although the ideal study design with parallel universes is not possible, thinking through such a study can provide insights that improve study design and analytic choices in real studies. Fundamentally, the question is whether the study participants are doing a good job representing the unobserved experience of the target population. For clarity, this article has focused on simple examples. However, there are more complex situations, such as time-varying confounding (9,10) and unfixable confounding that can occur and should be considered when evaluating confounding [see companion paper (1)]. The ability to minimize confounding through study design and analytic approaches depends upon a thoughtful approach to identifying sources of confounding because adjusting for factors that are not confounders can introduce bias instead of removing it. In the companion paper (1), we describe methods to assess sources of confounding.

Acknowledgments

I thank Olga Basso, Carey Drews-Botsch, Melanie Jacobson, Candice Johnson and Lisa Shandley for providing feedback on an earlier version of this manuscript. I thank the Whiteley Center for providing a peaceful place to write.

Funding

This study has not been funded by any external sources.

References

1. Howards PP. An overview of confounding. Part 2: how to identify it and special situations. Acta Obstet Gynecol Scand. 2018;97:400–406.
2. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology, 3rd edn. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008.
3. Makrides M, Crosby DD, Bain E, Crowther CA. Magnesium supplementation in pregnancy. Cochrane Database Syst Rev. 2014;4:CD000937. https://doi.org/10.1002/14651858.CD000937.pub2.
4. Maldonado G, Greenland S. Estimating causal effects. Int J Epidemiol. 2002;31:422–9.
5. Hernan MA. A definition of causal effect for epidemiological research. J Epidemiol Community Health. 2004;58:265–71.
6. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. Int J Epidemiol. 1986;15:413–9.
7. Mansournia MA, Higgins JP, Sterne JA, Hernan MA. Biases in randomized trials: a conversation between trialists and epidemiologists. Epidemiology. 2017;28:54–9.
8. Hernan MA, Robins JM. Estimating causal effects from epidemiological data. J Epidemiol Community Health. 2006;60:578–86.
9. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000;11:550–60.
10. Howards PP, Schisterman EF, Heagerty PJ. Potential confounding by exposure history and prior outcomes: an example from perinatal epidemiology. Epidemiology. 2007;18:544–51.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. A numeric example illustrating confounding and ways to adjust for it where maternal age is the only confounder.

Appendix S2. A numeric example illustrating confounding and ways to adjust for it where maternal age is a measured confounder but there is also an unmeasured confounder.

Appendix S3. Additional examples of research questions and sources of confounding.