

WIKIPEDIA

# Confounding

In statistics, a **confounder** (also **confounding variable** or **confounding factor**) is a variable that influences both the dependent variable and independent variable causing a spurious association. Confounding is a causal concept, and as such, cannot be described in terms of correlations or associations.<sup>[1][2][3]</sup>

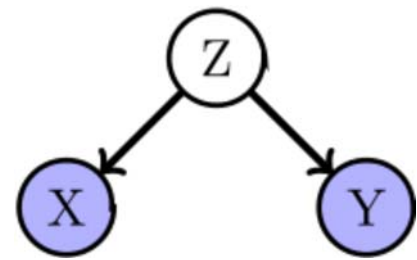


Illustration of a simple confounder. In other words Z is the cause of X and Y.

## Contents

**Definition**

**Control**

**History**

**Types**

**Examples**

**Decreasing the potential for confounding**

**See also**

**References**

**Further reading**

**External links**

## Definition

Confounding is defined in terms of the data generating model (as in the Figure above). Let  $X$  be some independent variable,  $Y$  some dependent variable. To estimate the effect of  $X$  on  $Y$ , the statistician must suppress the effects of extraneous variables that influence both  $X$  and  $Y$ . We say that,  $X$  and  $Y$  are confounded by some other variable  $Z$  whenever  $Z$  is a cause of both  $X$  and  $Y$ .

Let  $P(y \mid \text{do}(x))$  be the probability of event  $Y = y$  under the hypothetical intervention  $X = x$ .  $X$  and  $Y$  are not confounded if and only if the following holds:

$$P(y \mid \text{do}(x)) = P(y \mid x)$$

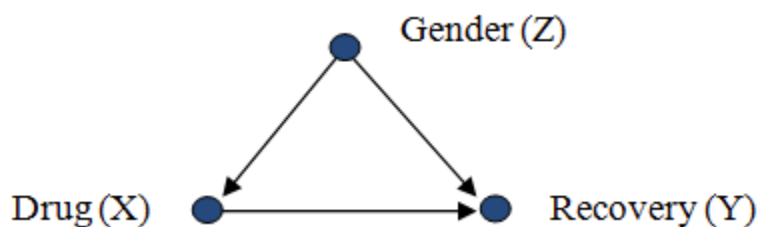
for all values  $X = x$  and  $Y = y$ , where  $P(y \mid x)$  is the conditional probability upon seeing  $X = x$ . Intuitively, this equality states that  $X$  and  $Y$  are not confounded whenever the observationally witnessed association between them is the same as the association that would be measured in a controlled experiment, with  $x$  randomized.

In principle, the defining equality  $P(y \mid \text{do}(x)) = P(y \mid x)$  can be verified from the data generating model assuming

we have all the equations and probabilities associated with the model. This is done by simulating an intervention  $\text{do}(X = x)$  (see [Bayesian network](#)) and checking whether the resulting probability of  $Y$  equals the conditional probability  $P(y | x)$ . It turns out, however, that graph structure alone is sufficient for verifying the equality  $P(y | \text{do}(x)) = P(y | x)$ .

## Control

Consider a researcher attempting to assess the effectiveness of drug  $X$ , from population data in which drug usage was a patient's choice. The data shows that gender ( $Z$ ) differences influence a patient's choice of drug as well as their chances of recovery ( $Y$ ). In this scenario, gender  $Z$  confounds the relation between  $X$  and  $Y$  since  $Z$  is a cause of both  $X$  and  $Y$ :



We have that

$$P(y | \text{do}(x)) \neq P(y | x)$$

because the observational quantity contains information about the correlation between  $X$  and  $Z$ , and the interventional quantity does not (since  $X$  is not correlated with  $Z$  in a randomized experiment). Clearly the statistician desires the unbiased estimate  $P(y | \text{do}(x))$ , but in cases where only observational data are available, an unbiased estimate can only be obtained by "adjusting" for all confounding factors, namely, conditioning on their various values and averaging the result. In the case of a single confounder  $Z$ , this leads to the "adjustment formula":

$$P(y | \text{do}(x)) = \sum_z P(y | x, z)P(z)$$

which gives an unbiased estimate for the causal effect of  $X$  on  $Y$ . The same adjustment formula works when there are multiple confounders except, in this case, the choice of a set  $Z$  of variables that would guarantee unbiased estimates must be done with caution. The criterion for a proper choice of variables is called the Back-Door <sup>[4][5]</sup> and requires that the chosen set  $Z$  "blocks" (or intercepts) every path from  $X$  to  $Y$  that ends with an arrow into  $X$ . Such sets are called "Back-Door admissible" and may include variables which are not common causes of  $X$  and  $Y$ , but merely proxies thereof.

Returning to the drug use example, since  $Z$  complies with the Back-Door requirement (i.e., it intercepts the one Back-Door path  $X \leftarrow Z \rightarrow Y$ ), the Back-Door adjustment formula is valid:

$$P(Y = \text{recovered} | \text{do}(x = \text{give drug})) = P(Y = \text{recovered} | X = \text{give drug}, Z = \text{male})P(Z = \text{male}) \\ + P(Y = \text{recovered} | X = \text{give drug}, Z = \text{female})P(Z = \text{female})$$

In this way the physician can predict the likely effect of administering the drug from observational studies in which

the conditional probabilities appearing on the right-hand side of the equation can be estimated by regression.

Contrary to common beliefs, adding covariates to the adjustment set  $Z$  can introduce bias. A typical counterexample occurs when  $Z$  is a common effect of  $X$  and  $Y$ ,<sup>[6]</sup> a case in which  $Z$  is not a confounder (i.e., the null set is Back-door admissible) and adjusting for  $Z$  would create bias known as "collider bias" or "Berkson's paradox."

In general, confounding can be controlled by adjustment if and only if there is a set of observed covariates that satisfies the Back-Door condition. Moreover, if  $Z$  is such a set, then the adjustment formula of Eq. (3) is valid <4,5>. Pearl's do-calculus provide additional conditions under which  $P(y|\text{do}(x))$  can be estimated, not necessarily by adjustment.<sup>[7]</sup>

## History

---

According to Morabia (2011),<sup>[8]</sup> the word derives from the Medieval Latin verb "confudere", which meant "mixing", and was probably chosen to represent the confusion between the cause one wishes to assess and other causes that may affect the outcome and thus confuse, or stand in the way of the desired assessment. Fisher used the word "confounding" in his 1935 book "The Design of Experiments"<sup>[9]</sup> to denote any source of error in his ideal of randomized experiment. According to Vandembroucke (2004)<sup>[10]</sup> it was Kish<sup>[11]</sup> who used the word "confounding" in the modern sense of the word, to mean "incomparability" of two or more groups (e.g., exposed and unexposed) in an observational study.

Formal conditions defining what makes certain groups "comparable" and others "incomparable" were later developed in epidemiology by Greenland and Robins (1986)<sup>[12]</sup> using the counterfactual language of Neyman (1935)<sup>[13]</sup> and Rubin (1974).<sup>[14]</sup> These were later supplemented by graphical criteria such as the Back-Door condition (Pearl 1993; Greenland, Pearl and Robins, 1999).<sup>[3][4]</sup>

Graphical criteria were shown to be formally equivalent to the counterfactual definition,<sup>[15]</sup> but more transparent to researchers relying on process models.

## Types

---

In the case of risk assessments evaluating the magnitude and nature of risk to human health, it is important to control for confounding to isolate the effect of a particular hazard such as a food additive, pesticide, or new drug. For prospective studies, it is difficult to recruit and screen for volunteers with the same background (age, diet, education, geography, etc.), and in historical studies, there can be similar variability. Due to the inability to control for variability of volunteers and human studies, confounding is a particular challenge. For these reasons, experiments offer a way to avoid most forms of confounding.

In some disciplines, confounding is categorized into different types. In epidemiology, one type is "confounding by indication",<sup>[16]</sup> which relates to confounding from observational studies. Because prognostic factors may influence treatment decisions (and bias estimates of treatment effects), controlling for known prognostic factors may reduce this problem, but it is always possible that a forgotten or unknown factor was not included or that factors interact complexly. Confounding by indication has been described as the most important limitation of observational studies. Randomized trials are not affected by confounding by indication due to random assignment.

Confounding variables may also be categorised according to their source. The choice of measurement instrument (operational confound), situational characteristics (procedural confound), or inter-individual differences (person

confound).

- An **operational confounding** can occur in both experimental and non-experimental research designs. This type of confounding occurs when a measure designed to assess a particular construct inadvertently measures something else as well.<sup>[17]</sup>
- A **procedural confounding** can occur in a laboratory experiment or a quasi-experiment. This type of confound occurs when the researcher mistakenly allows another variable to change along with the manipulated independent variable.<sup>[17]</sup>
- A **person confounding** occurs when two or more groups of units are analyzed together (e.g., workers from different occupations), despite varying according to one or more other (observed or unobserved) characteristics (e.g., gender).<sup>[18]</sup>

## Examples

---

In another concrete example, say one is studying the relation between birth order (1st child, 2nd child, etc.) and the presence of Down's Syndrome in the child. In this scenario, maternal age would be a confounding variable:

1. Higher maternal age is directly associated with Down's Syndrome in the child
2. Higher maternal age is directly associated with Down's Syndrome, regardless of birth order (a mother having her 1st vs 3rd child at age 50 confers the same risk)
3. Maternal age is directly associated with birth order (the 2nd child, except in the case of twins, is born when the mother is older than she was for the birth of the 1st child)
4. Maternal age is not a consequence of birth order (having a 2nd child does not change the mother's age)

In risk assessments, factors such as age, gender, and educational levels often affect health status and so should be controlled. Beyond these factors, researchers may not consider or have access to data on other causal factors. An example is on the study of smoking tobacco on human health. Smoking, drinking alcohol, and diet are lifestyle activities that are related. A risk assessment that looks at the effects of smoking but does not control for alcohol consumption or diet may overestimate the risk of smoking.<sup>[19]</sup> Smoking and confounding are reviewed in occupational risk assessments such as the safety of coal mining.<sup>[20]</sup> When there is not a large sample population of non-smokers or non-drinkers in a particular occupation, the risk assessment may be biased towards finding a negative effect on health.

## Decreasing the potential for confounding

---

A reduction in the potential for the occurrence and effect of confounding factors can be obtained by increasing the types and numbers of comparisons performed in an analysis. If measures or manipulations of core constructs are confounded (i.e. operational or procedural confounds exist), subgroup analysis may not reveal problems in the analysis. Additionally, increasing the number of comparisons can create other problems (see multiple comparisons).

Peer review is a process that can assist in reducing instances of confounding, either before study implementation or after analysis has occurred. Peer review relies on collective expertise within a discipline to identify potential weaknesses in study design and analysis, including ways in which results may depend on confounding. Similarly, replication can test for the robustness of findings from one study under alternative study conditions or alternative analyses (e.g., controlling for potential confounds not identified in the initial study).

Confounding effects may be less likely to occur and act similarly at multiple times and locations. In selecting study sites, the environment can be characterized in detail at the study sites to ensure sites are ecologically similar and

therefore less likely to have confounding variables. Lastly, the relationship between the environmental variables that possibly confound the analysis and the measured parameters can be studied. The information pertaining to environmental variables can then be used in site-specific models to identify residual variance that may be due to real effects.<sup>[21]</sup>

Depending on the type of study design in place, there are various ways to modify that design to actively exclude or control confounding variables:<sup>[22]</sup>

- **Case-control studies** assign confounders to both groups, cases and controls, equally. For example, if somebody wanted to study the cause of myocardial infarct and thinks that the age is a probable confounding variable, each 67-year-old infarct patient will be matched with a healthy 67-year-old "control" person. In case-control studies, matched variables most often are the age and sex. Drawback: Case-control studies are feasible only when it is easy to find controls, *i.e.* persons whose status vis-à-vis all known potential confounding factors is the same as that of the case's patient: Suppose a case-control study attempts to find the cause of a given disease in a person who is 1) 45 years old, 2) African-American, 3) from Alaska, 4) an avid football player, 5) vegetarian, and 6) working in education. A theoretically perfect control would be a person who, in addition to not having the disease being investigated, matches all these characteristics and has no diseases that the patient does not also have—but finding such a control would be an enormous task.
- **Cohort studies**: A degree of matching is also possible and it is often done by only admitting certain age groups or a certain sex into the study population, creating a cohort of people who share similar characteristics and thus all cohorts are comparable in regard to the possible confounding variable. For example, if age and sex are thought to be confounders, only 40 to 50 years old males would be involved in a cohort study that would assess the myocardial infarct risk in cohorts that either are physically active or inactive. Drawback: In cohort studies, the overexclusion of input data may lead researchers to define too narrowly the set of similarly situated persons for whom they claim the study to be useful, such that other persons to whom the causal relationship does in fact apply may lose the opportunity to benefit from the study's recommendations. Similarly, "over-stratification" of input data within a study may reduce the sample size in a given stratum to the point where generalizations drawn by observing the members of that stratum alone are not statistically significant.
- **Double blinding**: conceals from the trial population and the observers the experiment group membership of the participants. By preventing the participants from knowing if they are receiving treatment or not, the placebo effect should be the same for the control and treatment groups. By preventing the observers from knowing of their membership, there should be no bias from researchers treating the groups differently or from interpreting the outcomes differently.
- **Randomized controlled trial**: A method where the study population is divided randomly in order to mitigate the chances of self-selection by participants or bias by the study designers. Before the experiment begins, the testers will assign the members of the participant pool to their groups (control, intervention, parallel), using a randomization process such as the use of a random number generator. For example, in a study on the effects of exercise, the conclusions would be less valid if participants were given a choice if they wanted to belong to the control group which would not exercise or the intervention group which would be willing to take part in an exercise program. The study would then capture other variables besides exercise, such as pre-experiment health levels and motivation to adopt healthy activities. From the observer's side, the experimenter may choose candidates who are more likely to show the results the study wants to see or may interpret subjective results (more energetic, positive attitude) in a way favorable to their desires.
- **Stratification**: As in the example above, physical activity is thought to be a behaviour that protects from myocardial infarct; and age is assumed to be a possible confounder. The data sampled is then stratified by age group – this means that the association between activity and infarct would be analyzed per each age group. If the different age groups (or age strata) yield much different risk ratios, age must be viewed as a confounding variable. There exist statistical tools, among them Mantel–Haenszel methods, that account for stratification of data sets.
- **Controlling for confounding by measuring the known confounders and including them as covariates is multivariable analysis such as regression analysis**. Multivariate analyses reveal much less information about the *strength* or *polarity* of the confounding variable than do stratification methods. For example, if multivariate analysis controls for antidepressant, and it does not stratify antidepressants for TCA and SSRI, then it will ignore that these two classes of antidepressant have *opposite* effects on myocardial infarction, and one is

much *stronger* than the other.

All these methods have their drawbacks:

1. The best available defense against the possibility of spurious results due to confounding is often to dispense with efforts at stratification and instead conduct a randomized study of a sufficiently large sample taken as a whole, such that all potential confounding variables (known and unknown) will be distributed by chance across all study groups and hence will be uncorrelated with the binary variable for inclusion/exclusion in any group.
2. Ethical considerations: In double blind and randomized controlled trials, participants are not aware that they are recipients of sham treatments and may be denied effective treatments.<sup>[23]</sup> There is resistance to randomized controlled trials in surgery because it is argued that patients only agree to invasive surgery (which carry real medical risks) under the understanding that they are receiving treatment. Although this is a very real ethical concern, it is not however a complete account of the situation. For surgeries that are currently being performed regularly, but for which we have no concrete evidence of a genuine effect, surely it is unethical to continue without conducting sham control studies? In such a circumstance, thousands if not millions of people are going to continue to be exposed to the very real risks of surgery yet these treatments may possibly offer no discernible benefit. It is only via the use of sham-surgery as controls that medical science can determine whether a surgical procedure is efficacious or not. Arguably then, given that there are known risks associated with medical operations, it is extremely unethical to allow unverified surgeries to be conducted *ad infinitum* into the future. Yes it is undeniable that there are risks to the research participants in placebo-controlled studies if they receive the sham treatment - but those who receive the supposed "treatment" are exposed to the same risks and possibly for no gain. It is very obvious that the potential benefits to society from desisting with useless surgeries could be immense (and these benefits especially accrue to future sufferers of the same condition who might otherwise receive ineffective interventions and be exposed to an unnecessary medical risks). These potential benefits cannot simply be ignored if we are going to act ethically. Placebo-controlled studies offer the greatest clarification of what works and what does not and the ethical concerns raised by the resisters can be minimized by informing *all* research participants at study intake that they "may be" assigned to placebo group. To allow patients to volunteer when fully informed in this manner is to allow them to choose to benevolently gift the rest of us an advancement in scientific knowledge. It is not at all ethically obvious that the sufferers of medical conditions should be denied the opportunity to act as philanthropists by being denied the right to participate in such research. Surely it is their informed choice?

## See also

---

- Anecdotal evidence
- Causal inference
- Epidemiological method
- Simpson's paradox

## References

---

1. Pearl, J., (2009). Simpson's Paradox, Confounding, and Collapsibility In *Causality: Models, Reasoning and Inference* (2nd ed.). New York : Cambridge University Press.
2. VanderWeele, T.J.; Shpitser, I. (2013). "On the definition of a confounder". *Annals of Statistics*. **41**: 196–220. doi:10.1214/12-aos1058 (https://doi.org/10.1214%2F12-aos1058).
3. Greenland, S.; Robins, J. M.; Pearl, J. (1999). "Confounding and Collapsibility in Causal Inference". *Statistical Science*. **14** (1): 29–46. doi:10.1214/ss/1009211805 (https://doi.org/10.1214%2Fss%2F1009211805).
4. Pearl, J., (1993). "Aspects of Graphical Models Connected With Causality," In *Proceedings of the 49th Session of the International Statistical Science Institute*, pp. 391 - 401.
5. Pearl, J. (2009). Causal Diagrams and the Identification of Causal Effects In *Causality: Models, Reasoning and Inference* (2nd ed.). New York, NY, USA: Cambridge University Press.

6. Lee, P. H. (2014). *Should We Adjust for a Confounder if Empirical and Theoretical Criteria Yield Contradictory Results? A Simulation Study*. *Sci Rep*. **4**. p. 6085. doi:10.1038/srep06085 (<https://doi.org/10.1038%2Fsrep06085>).
7. Shpitser, I.; Pearl, J. (2008). "Complete identification methods for the causal hierarchy". *The Journal of Machine Learning Research*. **9**: 1941–1979.
8. Morabia, A (2011). "History of the modern epidemiological concept of confounding". *Journal of epidemiology and community health*. **65** (4): 297–300. doi:10.1136/jech.2010.112565 (<https://doi.org/10.1136%2Fjech.2010.112565>).
9. Fisher, R. A. (1935). The design of experiments (pp. 114-145).
10. Vandembroucke, J. P. (2004). "The history of confounding". *Soz Präventiv Med*. **47** (4): 216–224.
11. Kish, L (1959). "Some statistical problems in research design". *Am Sociol*. **26**: 328–338.
12. Greenland, S.; Robins, J. M. (1986). "Identifiability, exchangeability, and epidemiological confounding". *International Journal of Epidemiology*. **15** (3): 413–419. doi:10.1093/ije/15.3.413 (<https://doi.org/10.1093%2Fije%2F15.3.413>).
13. Neyman, J., with cooperation of K. Iwaskiewics and St. Kolodziejczyk (1935). Statistical problems in agricultural experimentation (with discussion). *Suppl J Roy Statist Soc Ser B* 2 107-180.
14. Rubin, D. B. (1974). "Estimating causal effects of treatments in randomized and nonrandomized studies". *J. Educational Psychology*. **66**: 688–701. doi:10.1037/h0037350 (<https://doi.org/10.1037%2Fh0037350>).
15. Pearl, J., (2009). *Causality: Models, Reasoning and Inference* (2nd ed.). New York, NY, USA: Cambridge University Press.
16. Johnston, S. C. (2001). *Identifying Confounding by Indication through Blinded Prospective Review*. *Am J Epidemiol*. **154**. pp. 276–284. doi:10.1093/aje/154.3.276 (<https://doi.org/10.1093%2Faje%2F154.3.276>).
17. Pelham, Brett (2006). *Conducting Research in Psychology*. Belmont: Wadsworth. ISBN 0-534-53294-2.
18. Steg, L.; Buunk, A. P.; Rothengatter, T. (2008). "Chapter 4". *Applied Social Psychology: Understanding and managing social problems*. Cambridge, UK: Cambridge University Press.
19. Tjønneland, Anne; Morten Grønbæk; Connie Stripp; Kim Overvad (January 1999). "Wine intake and diet in a random sample of 48763 Danish men and women". *American Society for Nutrition American Journal of Clinical Nutrition*. **69** (1): 49–54.
20. Axelson, O (1989). "Confounding from smoking in occupational epidemiology". *British Journal of Industrial Medicine*. **46**: 505–07. doi:10.1136/oem.46.8.505 (<https://doi.org/10.1136%2Foem.46.8.505>).
21. Calow, Peter P. (2009) *Handbook of Environmental Risk Assessment and Management*, Wiley
22. Mayrent, Sherry L (1987). *Epidemiology in Medicine*. Lippincott Williams & Wilkins. ISBN 0-316-35636-0.
23. Emanuel, Ezekiel J; Miller, Franklin G (Sep 20, 2001). "The Ethics of Placebo-Controlled Trials—A Middle Ground". *New England Journal of Medicine*. **345** (12): 915–9. doi:10.1056/nejm200109203451211 (<https://doi.org/10.1056%2Fnejm200109203451211>). PMID 11565527 (<https://www.ncbi.nlm.nih.gov/pubmed/11565527>).

## Further reading

---

- Pearl, J. (January 1998). "Why there is no statistical test for confounding, why many think there is, and why they are almost right" ([http://ftp.cs.ucla.edu/pub/stat\\_ser/R256.pdf](http://ftp.cs.ucla.edu/pub/stat_ser/R256.pdf)) (PDF). *UCLA Computer Science Department, Technical Report R-256*.

This textbook has a nice overview of confounding factors and how to account for them in design of experiments:

- Montgomery, D. C. (2001). "Blocking and Confounding in the 2<sup>k</sup> Factorial Design". *Design and Analysis of Experiments* (Fifth ed.). Wiley. pp. 287–302.

## External links

---

These sites contain descriptions or examples of confounding variables:

- [Linear Regression \(Yale University\) \(http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm\)](http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm)
  - [Tutorial by University of New England \(http://www.une.edu.au/WebStat/unit\\_materials/c1\\_behavioural\\_science\\_research/confounds.html\)](http://www.une.edu.au/WebStat/unit_materials/c1_behavioural_science_research/confounds.html)
- 

Retrieved from "<https://en.wikipedia.org/w/index.php?title=Confounding&oldid=820860025>"

---

**This page was last edited on 17 January 2018, at 01:12.**

Text is available under the [Creative Commons Attribution-ShareAlike License](#); additional terms may apply. By using this site, you agree to the [Terms of Use](#) and [Privacy Policy](#). Wikipedia® is a registered trademark of the [Wikimedia Foundation, Inc.](#), a non-profit organization.