Search...

Methodology

An introduction to research methods

+ Research approaches
+ Research designs
- **Variables and hypotheses**
  - Types of variables
  - Operationalization
  - Independent vs dependent
  - Mediator vs moderator
  - Confounding variables
  - Control variables
  - Writing hypotheses
+ Reliability and validity

Understanding confounding variables

Published on May 29, 2020 by Lauren Thomas. Revised on March 8, 2021.

In research that investigates a potential cause-and-effect relationship, a **confounding variable** is an unmeasured third variable that influences both the supposed cause and the supposed effect.

It's important to consider potential confounding variables and account for them in your research design to ensure your results are valid.

Table of contents

What is a confounding variable?

+ Sampling
+ Collecting data
+ Analyzing data
+ Writing up methods

Interesting topics

APA Style 6th edition
Applying to graduate school
Statistics
Chicago Style
Language rules
Methodology

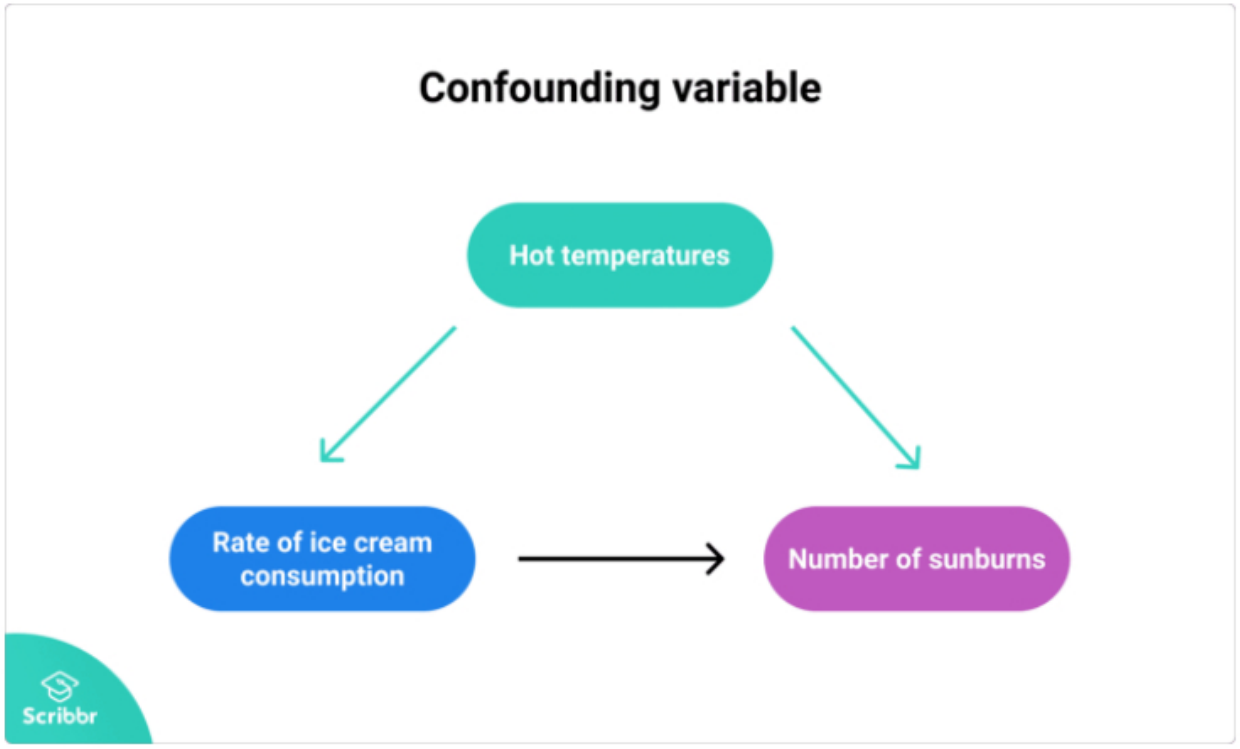
Confounding variables, which are also called confounders or confounding factors, are closely related to a study's independent and dependent variables. A variable must meet two conditions to be a confounder:

- It must be correlated with the independent variable. This may be a causal relationship, but it does not have to be.
- It must be causally related to the dependent variable.

Example of a confounding variable

You collect data on sunburns and ice cream consumption. You find that higher ice cream consumption is associated with a higher probability of sunburn. Does that mean ice cream consumption causes sunburn?

Here, the confounding variable is temperature: hot temperatures cause people to both eat more ice cream and spend more time outdoors under the sun, resulting in more sunburns.



Why confounding variables matter

To ensure the internal validity of your research, you must account for confounding variables. If you fail to do so, your results may not reflect the actual relationship between the variables that you are interested in..

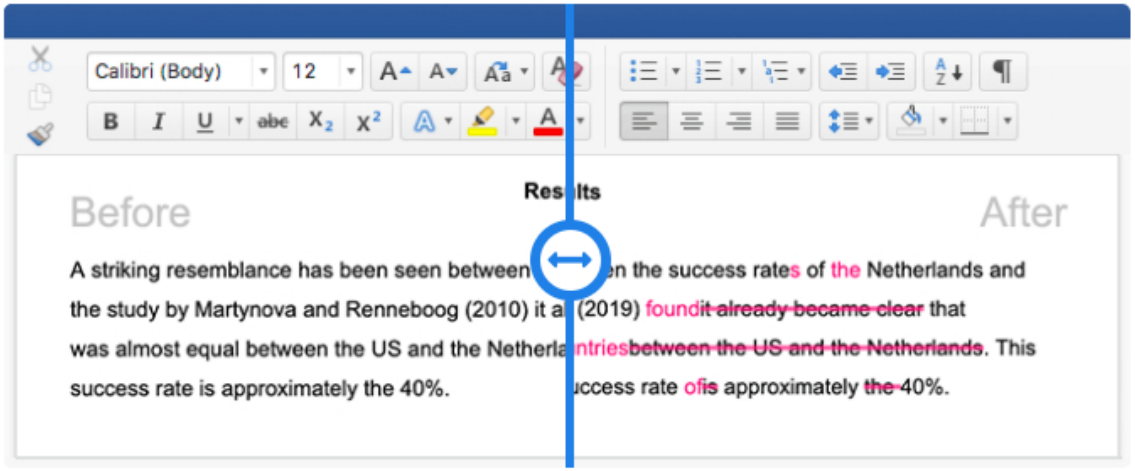
For instance, you may find a cause-and-effect relationship that does not actually exist, because the effect you measure is caused by the confounding variable (and not by your independent variable).

> **Example**
>
> You find that more workers are employed in states with higher minimum wages. Does this mean that higher minimum wages lead to higher employment rates?
>
> Not necessarily. Perhaps states with better job markets are more likely to raise their minimum wages, rather than the other way around. You must consider the prior employment trends in your analysis of the impact of the minimum wage on employment, or you might find a causal relationship where none exists.

Even if you correctly identify a cause-and-effect relationship, confounding variables can result in over- or underestimating the impact of your independent variable on your dependent variable.

> **Example**
>
> You find that babies born to mothers who smoked during their pregnancies weigh significantly less than those born to non-smoking mothers. However, if you do not account for the fact that smokers are more likely to engage in other unhealthy behaviors, such as drinking or eating less healthy foods, then you might overestimate the relationship between smoking and low birth weight.

writing by making sure your paper is free of vague language, redundant words and awkward phrasing.



See editing example

How to reduce the impact of confounding variables

There are several methods of accounting for confounding variables. You can use the following methods when studying any type of subjects—humans, animals, plants, chemicals, etc. Each method has its own advantages and disadvantages.

Restriction

In this method, you restrict your treatment group by only including subjects with the same values of potential confounding factors.

Since these values do not differ among the subjects of your study, they cannot correlate with your independent variable and thus cannot confound the cause-and-effect relationship you are studying.

Restriction example

You want to study whether a low-carb diet can cause weight loss. Since you know that age, sex, level of education and exercise intensity are all factors that may be associated with weight loss, as well as with the diet your subjects choose to follow, you choose to restrict your subject pool to 45-year-old

women with bachelor's degrees who exercise at moderate levels of intensity between 100–150 minutes per week.

- ✔ Relatively easy to implement
- ✖ Restricts your sample a great deal
- ✖ You might fail to consider other potential confounders

Matching

In this method, you select a comparison group that matches with the treatment group. Each member of the comparison group should have a counterpart in the treatment group with the same values of potential confounders, but different independent variable values.

This allows you to eliminate the possibility that differences in confounding variables cause the variation in outcomes between the treatment and comparison group. If you have accounted for any potential confounders, you can thus conclude that the difference in the independent variable must be the cause of the variation in the dependent variable.

Matching example

In your study on low-carb diet and weight loss, you match up your subjects on age, sex, level of education and exercise intensity. This allows you to include a wider range of subjects: your treatment group includes men and women of different ages with a variety of education levels.

Each subject on a low-carb diet is matched with another subject with the same characteristics who is not on the diet. So for every 40-year-old highly educated man who follows a low-carb diet, you find another 40-year-old highly educated man who does not, to compare the weight loss between the two subjects. You do the same for all the other subjects in your treatment sample.

- ✔ Allows you to include more subjects than restriction
- ✖ Can prove difficult to implement since you need pairs of subjects that match on every potential confounding variable
- ✖ Other variables that you cannot match on might also be confounding variables

Statistical control

If you have already collected the data, you can include the possible confounders as control variables in your regression models; in this way, you will control for the impact of the confounding variable.

Any effect that the potential confounding variable has on the dependent variable will show up in the results of the regression and allow you to separate the impact of the independent variable.

> **Statistical control example**
>
> After collecting data about weight loss and low-carb diets from a range of participants, in your regression model, you include exercise levels, education, age, and sex as control variables, along with the type of diet each subjects follows as the independent variable. This allows you to separate the impact of diet chosen from the influence of these other four variables on weight loss in your regression.

- ✔ Easy to implement
- ✔ Can be performed after data collection
- ✖ You can only control for variables that you observe directly, but other confounding variables you have not accounted for might remain

Randomization

Another way to minimize the impact of confounding variables is to randomize the values of your independent variable. For instance, if some of your participants are assigned to a treatment group while others are in a control group, you can randomly assign participants to each group.

Randomization ensures that with a sufficiently large sample, all potential confounding variables—even those you cannot directly observe in your study—will have the same average value between different groups. Since these variables do not differ by group assignment, they cannot correlate with your independent variable and thus cannot confound your study.

Since this method allows you to account for all potential confounding variables, which is nearly impossible to do otherwise, it is often considered to be the best way to reduce the impact of confounding variables.

Randomization example

You gather a large group of subjects to participate in your study on weight loss. You randomly select half of them to follow a low-carb diet and the other half to continue their normal eating habits.

Randomization guarantees that both your treatment (the low-carb-diet group) as well as your control group will have not only the same average age, education and exercise levels, but also the same average values on other characteristics that you haven't measured as well.

- ✔ Allows you to account for all possible confounding variables, including ones that you may not observe directly
- ✔ Considered the best method for minimizing the impact of confounding variables
- ✖ Most difficult to carry out
- ✖ Must be implemented prior to beginning data collection
- ✖ You must ensure that only those in the treatment (and not control) group receive the treatment

Frequently asked questions about confounding variables

What is a confounding variable? ›

What is the difference between confounding variables, independent variables and dependent variables? ›

Why do confounding variables matter for my research? ›

How do I prevent confounding variables from interfering with my research? ›

Is this article helpful? 122 14 0

Lauren Thomas

Lauren has a bachelor's degree in Economics and Political Science and is currently finishing up a master's in Economics. She is always on the move, having lived in five cities in both the US and France, and is happy to have a job that will follow her wherever she goes.

Other students also liked

Independent and dependent variables

In scientific research, the independent variable is the cause of a change in or effect on the dependent variable.

751

Understanding internal validity

Internal validity describes the extent to which a cause-and-effect relationship established in a study cannot be explained by other factors.

194

Control variables explained

A control variable is anything that is held constant in a study to prevent it from interfering with the results.

New

1 comment

Lauren Thomas (Scribbr Team) · May 29, 2020 at 2:15 PM

Thanks for reading! Hope you found this article helpful. If anything is still unclear, or if you didn't find what you were looking for here, leave a comment and we'll see if we can help.

Still have questions?

Name

Email

Let us know what's missing from this article

I'm not a robot

reCAPTCHA
Privacy - Terms

Add comment