# Confounding

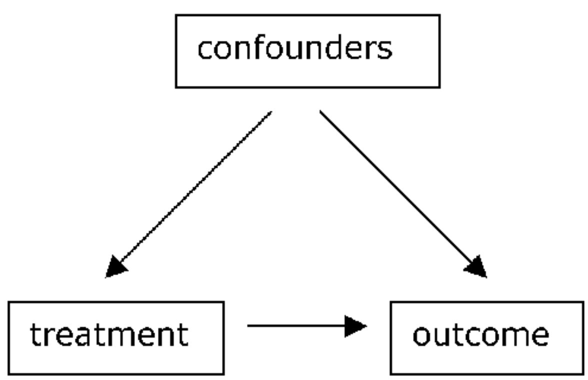
Reda Arab

## 1 Idea

If we want to assess a causal effect (*an exposure/treatment which causes an outcome*), we have to be careful about *confounding.*

In the Causal Inference framework, we use the term *confounding* to deal with a variable which has an effect both on the treatment/the exposure and on the outcome that we are observing. We can also see it as a variable which has an impact on the way we assign the treatment and on the outcome.

*Why ?* : It can lead to draw false conclusions as we will describe in the two following examples.

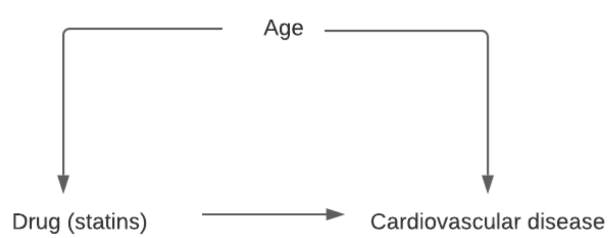


## 2 Exemple 1 : Effect of a medicine of cardiovascular diseases

Suppose that we want to assess the effect of a medicine (for example statins) on cardiovascular diseases. In this example, the exposure/treatment is taking the drug or no and the outcome is having a cardiovascular disease or no (e.g. we observe if it happens for one year after having taken the drug).

Suppose also that older people are more likely to take this drug than younger people and, at the same time, older people are more likely to develop cardiovascular diseases.

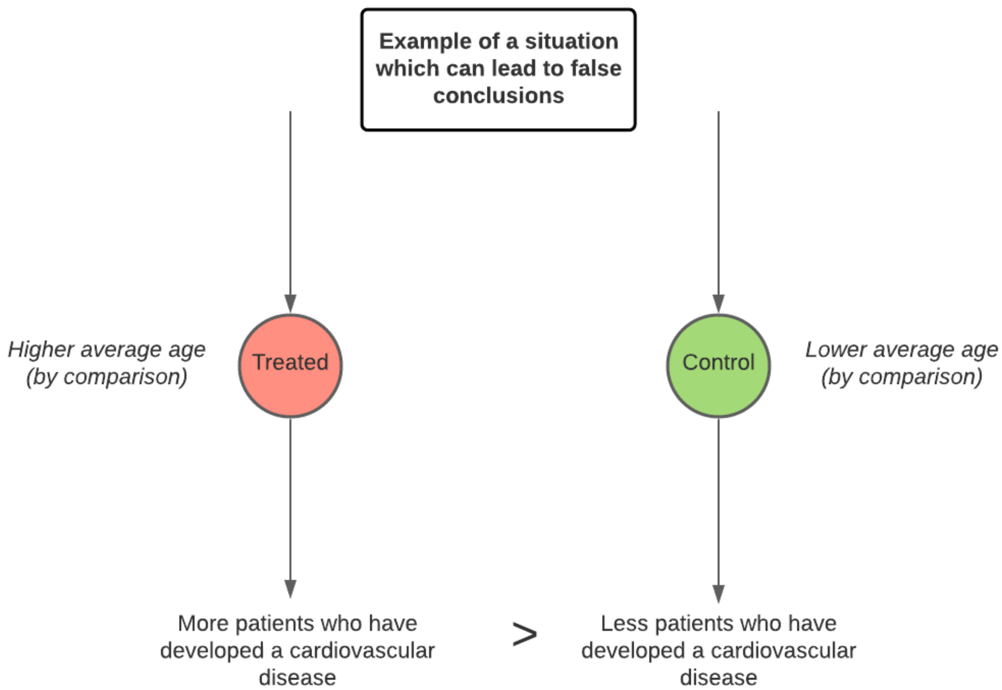
Our confounding factor here is *age*.




*What is the problem ?*

The problem is the following : to know whether there is an effect of this drug on cardiovascular diseases, someone might compare the number of persons who has developed a cardiovascular disease for one year in the treated group (the ones who take the drug) vs the control group (the ones who do not take it).

Thereby, we may have the treated group with an average age signficantly higher than the average age of the control group, given that older people are more likely to take the drug than younger people. Moreover, older people are more likely to develop a cardiovascular during the period of observation.

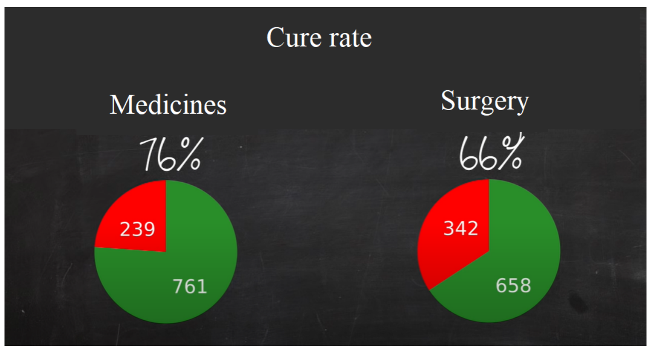



*Conclusion* : We might conclude that statins do not have effect (or even that it is harmful) given that we can find more people who have developed a cardiovascular disease within the treated group (even if there is the same number of people in the two groups). This first example shows that taking into account confounding is crucial.

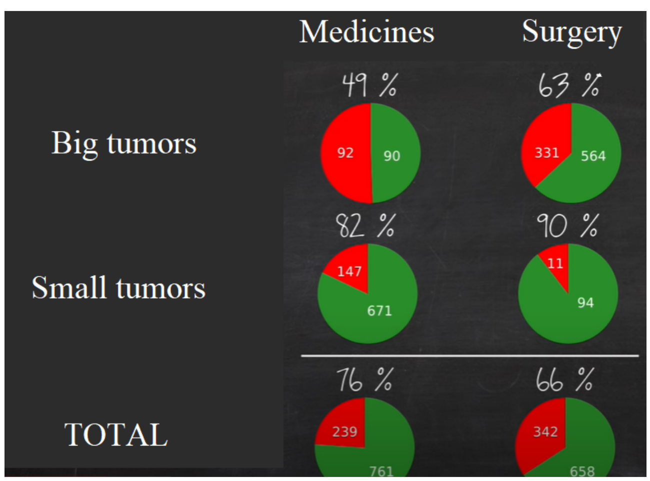
# 3 Exemple 2 : Effect of an intervention(medicine/surgery) on a cure rate

This example and the charts come from a video of ScienceEtonnante (French physicist and Youtuber) and you can have access to the video here : `https://www.youtube.com/watch?v=vs_Zzf_vL2I`.

Suppose that we have two treatments known to cure a certain type of tumors : medicines and surgery. We want to know what is the best one and to know what is the effect of a treatment on the cure rate.




By looking at these first charts, we might conclude that medicines are more effective than surgery. Let's analyze our data more carefully.

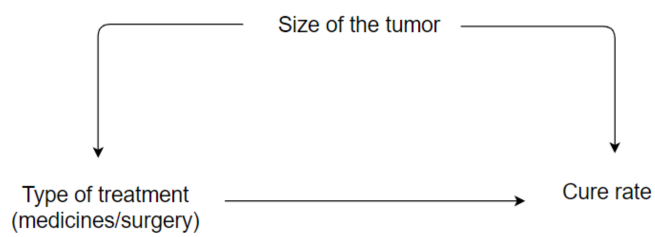



By splitting into two groups according to tumors sizes, we can observe that surgery is the best treatment in each case.

*What is the problem ?*

Surgery is largely more used when doctors have to deal with bigger tumors (as it can be seen on the charts). Moreover, bigger tumors imply a cure rate which is lower than the cure rate for smaller tumors.

Our confounding factor in our example is *size of the tumor* which affects both the treatment (which treatment we use) and the outcome (cure rate).




*Conclusion :* By analyazing our data without taking into account this confounding factor, it can lead to false conclusions (saying that taking medicines is more effective than surgery). Once again, someone interested in assessing causal effects should be careful about potential confounding factors.