**[On chocolate and Nobel prizes: Correlation is not causation](http://chronicle.umbmentoring.org/on-methods-correlation-is-not-causation/%22%20%5Co%20%22Permanent%20Link%20to%20On%20chocolate%20and%20Nobel%20prizes%3A%20Correlation%20is%20not%20causation)**

 Editors Note: Last month, Adar and I saw a wonderful  chart in the *New England Journal of Medicine*, showing the association between per capita chocolate consumption and the number nobel prizes per 10 million population (see below). Since chocolate consumption is unlikely to *cause* Nobel prizes, this association should be considered correlation which is accounted for by other, unmeasured variables (e.g., SES, investment in science). This led to a discussion about common misperceptions regarding cause and effect and Adar’s informative new post on the topic. Enjoy!

“Circumstantial evidence is a very tricky thing. It may seem to point very straight to one thing, but if you shift your own point of view a little, you may find it pointing in an equally uncompromising manner to something entirely different.”A.C. Doyle

**How is correlation different from causation?**by Adar Ben-Eliyahu, Ph.D.

I[n a recent post](http://chronicle.umbmentoring.org/evidence-corner-is-mentoring-worth-the-investment-the-jury-is-out/), Dr. David DuBois notes that “the jury is out” when it comes to mentoring. Closely related to the need to substantiate the dollar value of mentoring, is the need to determine its effectiveness. Most of the statistics that are used to determine the association between mentoring and various outcomes are based on correlations. Below, I explain correlational analysis and, how it is different from causation – or the conclusion that one thing leads to another.



 **What is a correlation?**

A correlation is the basic analyses on which quantitative analyses are built. A correlation simply means that there is a relation or association between two things.

There are three types of correlational relations:

1)   *Positive correlation:* As one value increases (i.e., mentoring quality), the other one increases (i.e., mentee satisfaction).

2)   *Negative correlation:* As one value increases (i.e., social support), the other values decreases (i.e., negative emotions).

3)   *Null or No correlation*: There is no association between the increase or decrease in values (i.e., height of the mentee and mentoring quality).

When we use statistical analyses, we try to determine whether relations between variables are related to each other versus having not (or null) correlation. If we cannot determine that in most cases (95% or more) there is a relationship (e.g., relationships quality and satisfaction), we would conclude that there is insufficient evidence to conclude that these two variables/constructs are related to each other.’

However, with a different sample, more participants, or slightly different measures, we might be able to find these relations. That is why it is important to pay close attention to details!

Below I present examples from the mentoring literature illustrating the positive and negative correlation.

**Positive Correlation**

A positive correlation exists when two variables tend to increase (or decrease) together. An example from the mentoring literature can be seen in [Phinney et al.’s findings about mentoring relationships and mentee satisfaction](http://chronicle.umbmentoring.org/exploring-a-mentoring-program-for-latino-college-freshmen).  They found that as mentoring quality increases, mentee satisfaction increases (see Table 2 on page 612).

This type of association can be illustrated as follows:

 

 Notice that as one value increases, so does the other.

**Negative Correlation**

A negative correlation means that as one variable or quality increases, another one decreases. An example from the mentoring literature can be seen in the recently summarized paper on supportive relationships and rejection sensitivity in adolescents by [McDonald et la., (2010)](http://chronicle.umbmentoring.org/supportive-relationships-and-rejection-sensitivity-in-adolescents/)

In their research, McDonald et al. found that social support is associated with lower levels of anxiety, anger and depressive symptoms. In general, this was true for social support from mother, father, and best friend.  Essentially, what this means is that there was a negative correlation between support and negative emotions. As support increases, negative emotions decrease (see Table 1 on page 567 of McDonald et al., 2010).

We call this a negative correlation because support is inversely related to negative emotions. This relation can be depicted in a figure in the following way:



 Notice, that as one value increases, the other value decreases.

Even if we find a positive or negative correlation that is significant (high probably that is not null or no correlation), we still need to think about whether there might be other reasons that the qualities or variables are related.

In this way, *correlation does not always mean that there are causal relations between variables or qualities.*

**What would be some reasons that a significant correlation is not showing us that one construct predicts another?**

Many statistical analyses are based on correlations. However, before we can say that one component/quality *predicts* another, we need to substantiate a number of criteria.

First, we need to make sure that the sequence of events makes sense. For example, in the McDonald et al study, it could be that youth who are less angry, less anxious, and less depressed, are more receptive to receiving social support. That is, it might not be the case that support is leading to reduced negative emotions, but rather youth who tend to have lower levels of negative emotions, tend to be more open and receptive to support from their parents and peers. In this way, correlations do not substantiate a causal relation between variable/qualities/constructs, but suggest that such a relation exists. We would need a longitudinal design, or to survey participants over a number of time points, in order to better determine whether supportive relationships (for younger children) causes less negative emotions (for youth). However, this does not mean that there is no value to these findings; we just have to be cautious about how we use them.

One way to make sure that the relation is causal, is by conducting longitudinal research design, whereby measurement of one variable is obtained before the other one. Phinney et al. did this when they surveyed mentor-mentee pairs in the fall about their relationship quality, and followed up in the spring about the mentee satisfaction. In this way, relationship quality in the fall was measured prior to the relationship and can be said to predict mentee satisfaction in the spring.

However, we cannot assume a causation of this sort for every finding. For example, surveying youth in high-school might produce the finding that height in the winter is related to IQ in the spring, simply because seniors tend to be taller than freshmen, and also have gained knowledge. In this way, it would be a third variable that leads to the relations between the two constructs: age/growth.

1) with age, one grows taller

AND

2) knowledge gains also increase with age

So really, age predicts height and knowledge. We therefore find an association such that as height increases, so does knowledge. However, it is because both height and knowledge increase due to the third variable – age/growth – that they seem to be related. Thus, age is a third variable that is accounting for the relationship between height and knowledge.

**Conclusion**

Not every significant association implies that one variable causes the other. The fact that there might be other explanations for why certain variables are associated, does not necessarily weaken the finding, it just means that we have to be cautious about conclusions that we draw. And, although chocolate may not cause you to win a Nobel prize, I hope you don’t let this disappointment ruin your appetite for statistical knowledge (or chocolate!).

<http://chronicle.umbmentoring.org/on-methods-correlation-is-not-causation/>