Correlation is not causation

By C. Carter-Snell, RN, MN

In a recent news release, a link was announced between increased antibiotic use and the development of breast cancer (CBC staff, 2004). While researchers were careful in the article to reinforce that cause and effect have not been established, this was not clear in subsequent television and radio reports about the research. All too often, we hear of "links" or "associations" between variables and begin to consider this proof that one causes the other. Many health research studies involve testing for associations between variables and are referred to as correlational studies. Patients may bring in articles and ask about the studies, or we may be looking for information to support our practices. Either way, we need to be familiar with the meaning of studies discussing associations or links between variables. The purpose of this article is to explore the concept of correlation, describe correlational studies, and then to discuss the criteria required for establishing causation.

Correlation

Correlation is the extent of a relationship which is found between two or more variables. If one variable always changes in direct relationship to another, this is referred to as a "perfect" correlation. Numerically, a perfect correlation is written as 1.0 in a statistical report. It may be -1.0 or +1.0 depending on the direction of the relationship. A positive correlation, or a value between 0 and +1.0, means that as the level of one variable increases, the other variable also increases. A negative correlation is also known as having an inverse relationship. This is expressed as ranging from 0 to -1.0 numerically and means that as one variable increases, the other variable decreases.

Correlations are sometimes described as being weak or strong. There is not an absolute value required for a strong or weak relationship, however. It depends on the type of variables and the implications of the relationship. For instance, if deciding between two types of blood pressure machines, a positive correlation of 0.70 is not very strong and you would not want to switch machines. In contrast, if looking at variables with less precise measures such as psychosocial measures of stress, or quality of life, a lower correlation would be acceptable. For this type of variable, a correlation of 0.70 would be fairly high (Loiselle, Profetto-McGrath, Polit & Beck, 2004). In Table One we see a strong positive relationship, with one variable increasing as the other increases. Note in Table One that the scores are represented in the table as dots on the scatterplot. The arrow represents the "regression line" of the data, or the slope of the line which is the best fit to the data.

The type of test used for correlation depends on the level of data being used. Spearman's rank order correlation (rho) is an example of a test used for association between variables which are nominal or ordinal in nature. Nominal variables are those in which the characteristic is either present or absent (e.g., obese or not obese). Ordinal variables have different levels of the characteristic, but there are not equal intervals between them. Examples would be thin, average, and obese, or the use of a scale from one to five (a Likert scale). This is a nonparametric form of correlation.

A common test used to determine correlation with more precise levels of data is Pearson's r or Pearson's correlation coefficient. This test is used for interval level data in which there are equal intervals between each value, such as the number of kilograms a person weighs. Pearson's r is also used for ratio level data, meaning there are equal intervals and an absolute zero or absence of the characteristic, such as the amount of weight gain or the amount of food eaten. It is important, though, that even at this level of precision, the data be "normally distributed", otherwise nonparametric correlation must be used (Greenhalgh, 1998). Normal distribution refers to ensuring there is a wide range of data, with the bulk of the data forming a bell-shape curve on a graph (otherwise known as a normal curve). A normal curve is difficult to obtain when subjects are not randomly selected or assigned to groups.

It is possible to look at the relationship between more than two variables. One such test is called multiple regression. Multiple regression has sometimes been explained as similar to correlation. There is a similarity with correlation in that both look at relationships between variables and both have regression lines, looking for the best fit of the data around this line. The difference with multiple regression is that there is a dependent or target variable identified. The remaining independent variables are then combined in different ways until the combination is found which best predicts the target variable. This implies causation, but it is still not established (Greenhalgh, 1998).

Study design and correlational studies

Research designs are of three main levels: experimental, survey, or exploratory-descriptive. These are summarized in Table Two. Note that correlational designs appear in the survey design category.

In order to understand why correlational designs cannot establish causation, we need to first explore the requirements for experimental designs. An experimental design has three major criteria (Greeno, 2002):

- there is random assignment of the subjects to one of two or more groups
- at least one of the key variables is manipulated or used in varying levels

- there is an attempt to control factors which may cause alternate explanations for the results (called confounding or extraneous variables)

Within the experimental design category there is sometimes further classification. A "true" experimental design is considered one in which there is both random selection of the subjects and random assignment to groups, in addition to the manipulation and control of extraneous variables. This level of design is most commonly found in laboratories and is difficult in health care research. If only random assignment to groups is used, and random selection is not possible, some researchers refer to this as a "Quasi-experimental" design. According to Greeno's categories, however, it would still be within the experimental category. The significance of this difference is in the risk for alternate explanations for the results. If the sample is randomly selected and randomly assigned, then it is assumed that the differences in individuals will be randomly distributed between the groups. If all the available subjects are used (a "convenience" sample) and they are only randomly assigned, you could end up with all the unusual cases in one group and confuse the interpretation of the results. An example is a research study on suctioning nosocomial pneumonia in endotracheally intubated patients (Carter-Snell & Sheehan, 1989). All patients who had recently been endotracheally intubated were included in the study (lack of random selection) but randomly assigned to one type of suctioning or the other. As it happened, all of the patients who vomited during intubation ended up in the same group and that group had a slightly higher pneumonia incidence. Instead of attributing this to the suctioning method, it well could have been the irritation and aspiration behind this difference.

In survey research, nothing can be manipulated or changed, but you may discover a relationship of interest. Examples of descriptive research designs include the case report, case-series report, cross-sectional studies, surveillance, and correlational studies (Grimes & Schulz, 2002). A correlational study design is used to explore the relationship between two or more variables when there is not very much data to support causal relationships (Brink & Wood, 1998). Data is obtained, usually from large samples in the field, and then variables are examined for relationships. There is no ability to randomly assign members to groups, nor is any variable manipulated. One of the largest and longest correlational studies is the Nurses' Health Study (Linton, 2004) at Harvard University. It has been running since 1974 with nurses sharing their health, fitness, medical, and personal variables. These have provided directions for key studies involving variables which may affect risk for heart disease, breast cancer, and many other diseases. This study has played a significant role in raising questions among researchers and being the drive behind subsequent randomized controlled trials.

Table One: Correlation table for food intake and weight

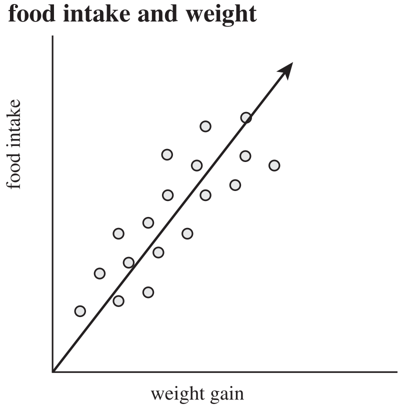

Table Two: Types of research designs	
Level of Design	**Type of Study**
I. **Exploratory-descriptive** • used to explore data or theories with little prior data	• Descriptive • Exploratory
II. **Survey designs** • Statistically analyze relationships	• Comparative • Correlational
III. **Experimental** • used to test theory	• Descriptive • Exploratory

In exploratory-descriptive research there is very little data available, or very few studies which explore a relationship. An example of the descriptive design is the census data. An exploratory design has the least controls and usually includes the use of qualitative methods to explore small samples indepth. Examples include ethnographic or grounded theory studies. Historical research is another example of exploratory-descriptive research.

Causation and correlational studies

Three major criteria have been described as necessary before inferring cause (Cook & Campbell, 1979). These include: covariation between cause and effect; the occurrence of the cause before the effect; and the use of control to rule out other possible causes for the effect. This notion of causality is consistent with the experimental design, not the correlational design. If one manipulates the cause, it should follow that the effect is also manipulated. According to this definition of causation, therefore, only experimental designs would be suited to establishing causation.

There are some who argue that correlational studies can in fact be used to determine causality as long as certain criteria are met. Examples of these were developed by the US Surgeon General in the 1960s and were subsequently updated (Reynolds, 1999):

- consistency - whether the association appears in multiple studies at similar levels
- strength - the size of the association
- specificity - the appearance of the association mainly when the causal variable is present
- temporality - occurrence of the probable causal variable before the association
- coherence - the ability to explain the association with known facts

Reynolds (1999) disputes the ability to use these criteria causally, citing the case of the over 100 correlational studies on smoking and low birthweight babies. He presents a review of the research in which it was found that mothers who smoke also have higher caffeine intake as well as increased likelihood of narcotics and other illicit drugs. Each of these could explain the low birthweights. As noted by Greeno (2002), giving up any of the conditions of the experimental design removes our ability to make causal inferences.

Conclusion

Although there are some who would like to use correlational studies to imply causation, we have seen here that these types of studies would not meet the criteria for causation. When we hear words like "linked to" or "associated with", we know that the researchers were using correlational research in some form. We then know to expect that there may be many other unknown factors which could be alternate explanations for the association. An often-told legend in statistics is used to illustrate the error in this thinking. Years ago, researchers were investigating the factors associated with the development of malaria. They discovered there was a strong association between the amount of rainfall in tropical zones and the incidence of malaria. They believed that the malaria parasites were carried in the rainfall. This was reinforced when researchers noted that those who wore protective clothing against the rain had a decreased incidence of malaria. This type of thinking excluded other possibilities which may explain why these two factors were linked. It was not until much later that they realized it was the Anopheles mosquitos that actually carried the parasites and transmitted the disease with their bite. The mosquito population would, of course, increase in times of high rainfall, and those who wore protective raingear would also be less likely to be bitten by mosquitos. Let's look beyond the rain in our interpretation of the research and try to find the mosquitos!

References

Brink, P.J., & Wood, M.J. (1998). Introduction (Chapter 1). In P.J. Brink, & M.J. Wood (Eds.), **Advanced design in nursing research** (2nd ed., pp. 3-20). Thousand Oaks: Sage.

Carter-Snell, C.J., & Sheehan, G. (1989). Comparison of two suction techniques for intubated ICU patients [Abstract]. **Proceedings of the Twenty-eighth Interscience Conference on Antimicrobial Agents and Chemotherapy** (p. 116).

CBC Staff. (2004). US Study suggests antibiotic, breast cancer link. **CBC News**. Retrieved February 20, 2004 from http://www.cbc.ca/stories/2004/02/16/antibiotic_bcancer040216

Cook, T.D., & Campbell, D.T. (1979). Causal inference and the language of experimentation. In T.D. Cook & D.T. Campbell (Eds.), **Quasi-experimentation: Design & analysis issues for field settings**. Boston: Houghton Mifflin Company.

Greenhalgh, T. (1998). Statistics for the nonstatistician - II: "Significant" relations and their pitfalls. **Student BMJ, 6**, 156-159.

Greeno, C.G. (2002). Major alternatives to the classic experimental design. **Family Process, 41**(4), 733-736.

Grimes, D.A., & Schulz, K.F. (2002). Descriptive studies: What they can and cannot do. **Lancet, 359**(9301), 145-149.

Linton, M. (2004, January 9). Nurses' health study a landmark. **C-Health, Toronto Sun.** Retrieved March 3, 2004 from http://www.canoe.ca/Health0109/04_linton-sun.html

Loiselle, C.G., Profetto-McGrath, J., Polit, D.F., & Beck, C.T. (2004). Analyzing quantitative data. In C.G. Loiselle, J. Profetto-McGrath, D.F. Polit, & C.T. Beck (Eds.), **Canadian Essentials of Nursing Research** (pp. 329-380). Philadelphia: Lippincott.

Reynolds, C. (1999). Inferring causality from relational data and designs: Historical and contemporary lessons for research and clinical practice. **The Clinical Neuropsychologist, 13**(4), 385-395.

'Brain death' in Vancouver!

By Caroline McGarry-Ross, RN, ENC(C), Halifax, NS

From April 9-11, 2003 Vancouver hosted a Canadian forum with one main goal: to bring experts from many fields together in order to develop a new Canadian protocol for assessing and diagnosing the patient with brain (versus cardiac) death. The forum was officially called "Severe Brain Injury to Neurological Determination of Death", and NENA was invited to send three emergency nurses from across Canada to participate. Provincial associations were contacted and requested to forward interested and appropriate names for selection. By some divine intervention, my name was selected, and so it was that I found myself attending this conference along with two other ER nurses: Clay Gillrie from British Columbia, and Francoise " Frankie" Verville from Saskatchewan.

Now one would think that five days in downtown Vancouver at the posh Fairmont Hotel Vancouver (all expenses paid by the conference) would be a treat. And, in many ways, it was. The catch was that I had to actually *participate* in this conference (versus attend and just look interested, which is something I can do well...) and it was on a subject that I had managed to avoid for years: how best to identify and diagnose the patient with brain death. You see, a couple of negative experiences in emergency and ICU had left me with a rather uncomfortable feeling about the whole process and that, coupled with a nagging spiritual concern for patients who became donors, simply meant avoiding such situations as much as possible. That is where the divine intervention comes in; clearly the good Lord decided it was high time I learned much more about this. And so, with some hesitation about it all, as well as Air Canada's ability to get me there, off to Vancouver I went.

On day one of the conference, we did the obligatory introductions and I found myself surrounded by quite an esteemed (and very cerebral) group from all over Canada. Indeed, there were neurosurgeons, neurologists,