



Dynamic Chiropractic – November 19, 1993, Vol. 11, Issue 24

Scientific Writing: Discriminating Good from Bad, Part III

Assignment, Assessment, and Analysis in Experimental Studies

By Arthur Croft, DC, MS, MPH, FACO

An experimental study or clinical trial is similar to the prospective study discussed in part II of this series except for an important difference: While patients self-assign themselves to cohort groups in the case of a prospective study, in a clinical trial the researchers assign participants into either the control group or the study group. This process must be entirely random so that the participants have an equal chance of being assigned to either group. As the name implies, this process is called "randomization." Lack of randomization is a very serious but common flaw in medical/chiropractic studies. A randomized clinical trial is the only truly acceptable form of clinical trial.¹

There have been few controlled, randomized, clinical trials of chiropractic treatment and we, as a profession, should consider ourselves somewhat derelict in that regard. Still, some of our greatest critics live in glass houses. I recall a dinner conversation I had with Alf Nachemson at a Canadian multidisciplinary conference several years ago. He was chiding our profession for its procrastination with such research. "Show it to me and then I'll believe you," he said, meaning, show me the results of controlled, randomized, blinded trials. Recently, however, Nachemson has stated that no trials exist which would lend scientific validity to most forms of spinal surgery² and he has called for a moratorium on unproven invasive maneuvers.³

Similarly, in the cervical spine, Rowland⁴ has called attention to the lack of clinical trials supporting the usefulness of surgical decompression in cases of cervical spondylotic myelopathy, noting that the entrepreneurial spinal surgeon has effectively circumnavigated this issue through the application of Catch-22 logic. The reasoning is that, since we know (or rather presume) that such conditions are progressive and will lead to greater degrees of neurological impairment, it would be grossly unethical to place patients at risk by assigning them to control groups which would deny them the necessary surgery. Rowland points out that many prospective studies have not supported this logic; in some studies those

choosing surgery have been no better off in the long run than those electing to forego the procedure.

Since so much of our literature is the result of clinical trials, let us develop our framework of analysis around them. Riegelman⁵ provides a good basis for this approach. He divides the analysis into five parts: 1) assignment; 2) assessment; 3) analysis; 4) interpretation; and 5) extrapolation. In this installment we will discuss numbers 1-3. In part IV of this series we will discuss numbers 4 and 5).

Assignment

Selection bias, a common problem in study design, occurs when researchers unintentionally introduce factors into a study which will bias the results. Let's look at an example. In a recent study designed to measure the long-term prognosis in cervical acceleration/deceleration (CAD) trauma, the researchers selected 100 consecutive cases from the files of a retired consultant (IME doctor). Since one of the goals of this paper was to measure the effect of litigation on the outcome, only cases which had been litigated were used. The authors found that 86 percent of these patients continued to be symptomatic an average of 13 years post-injury and concluded that litigation had no effect on outcome.

The first question we must ask is, "Were these patients representative of all patients sustaining CAD injury?" We must assume that some persons are not injured in minor accidents. Yet this group is not represented in our present study. It is conceivable that a more seriously injured group would bring suit against the other party and would therefore be seen by this consultant. This example illustrates two fundamental concerns with selection bias: The study population was probably not representative of all persons exposed to CAD, and this difference probably did effect the results.

As I mentioned in an earlier part of this series, study and control groups must be matched as closely as possible to avoid the possible influence of confounding variables. Problems such as this can be avoided by careful study design. If we were going to study the incidence of respiratory disorders in coal miners and did not take into consideration the smoking habits of these individuals, smoking would become an important confounding variable. To account for cigarette smoking as an independent factor, we would separate smokers and nonsmokers into two cohorts.

Of the more obvious examples of confounding variables are age and gender, and yet subtle variables related to gender are sometimes overlooked. In one study, designed to compare the rate of lung cancer with smoking, researchers found a slightly higher correlation between smoking and the development of lung

cancer in women. In the male population of smokers were a large number of cigar and pipe smokers who (like our president) did not inhale and were thus less at risk for lung cancer. Women, on the other hand, rarely smoke anything but cigarettes (unlike our president), and when researchers adjusted for this and compared only those men and women who smoked cigarettes, the gender variance disappeared.

Assessment

Before assessment of a study, researchers must define the "outcome" or expected results. Difficulties in this area are common and can create serious flaws in a project. The most common problems to look for: 1) authors don't define an outcome; 2) measurement of the outcome is inappropriate (as using plain film radiography to determine spinal stenosis); 3) measurement of outcome is not precise (asking patients to rate their pain as "mild, moderate or severe" rather than using the more acceptable visual analog pain scale); 4) measurement of outcome is incomplete (as when a number of participants drop out of the study or when some patients are measured at different times from others. An example of this would be to assess the levels of pain in some patients on their days off and in others after a full day of work); 5) outcome is affected by the process of measuring, e.g., repeatedly measuring the cholesterol levels in a study group may consciously or subconsciously cause them to alter their diets/lifestyles. A classic example is the long-term smoking study where researchers follow two cohorts, one composed of smokers; the other, nonsmokers. Over time, a number of the smokers quit smoking and essentially switched cohorts.

Analysis

The analysis of data in clinical trials revolves around significance testing. To begin this process we must first state an hypothesis. For example, in a controlled trial of spinal manipulation, we might hypothesize that patients with acute torticollis will respond sooner when treated by a chiropractor than those who merely take ibuprofen. We then formulate a null hypothesis which says exactly the opposite, i.e., chiropractic care does not provide any measurable advantage over ibuprofen. Later, if we find there is a significant effect, we will claim to have rejected the null hypothesis.

Next we must decide the level of significance or probability we wish to test for. In most research, an arbitrary figure of five percent is chosen. Usually it is expressed in the familiar form, $p < 0.05$, which simply means that if, in our example, the group treated with chiropractic care did have a better outcome than the other group, the probability that this difference is the result of chance alone is less than five percent. When the results of research are more critical, researchers may choose a one percent level of significance.

Unfortunately this p value analysis can often be misused, misinterpreted or misunderstood. We often hear the statement, "the results failed to reach significance." This can be misleading because if we arbitrarily set our probability level at $p < 0.05$ and our results come in at $p = 0.06$, we will have to conclude that the results are not significant. However there is still only a six percent chance that the observed effect was due to chance alone. Perhaps, on reflection, we might accept that level of uncertainty.

Also, although I will spare you the calculation, the smaller the sample size the less significant the results will be. If we tossed a coin four times and got one heads and three tails we would consider that variance to be merely the result of chance and therefore not very significant. But, if we were to toss the same coin 400 times and came up with 100 heads and 300 tails we would be forced to conclude that something other than chance was at play and therefore the results are probably significant. In fact, the larger the sample size, the more likely we are to find significant correlations, while the smaller the sample size, the more likely it is that we will miss a significant correlation.

In the future you will see a gradual phasing out of p value reporting. Replacing it will be the value, "confidence interval" or CI. This will help eliminate some of the current confusion regarding the "significance" of research results.

Once we have gathered our data, we subject that data to the appropriate statistical significance test. The one we use will depend on the type of data we have collected, the size of the study, the number of independent variables, the number of dependent variables, whether there are any co-variables, and so on (reference number one is an excellent source for more information in this area). The test to be used is always an important consideration in the original study design.

It is possible to falsely reject or falsely fail to reject the null hypothesis. The former example is referred to as a type I error. The latter is a type II error. And, just as small sample sizes may result in type II errors, very large sample sizes may result in type I errors. One must remember that not all statistically significant differences are important, and not all important differences are statistically significant.

In a recent study which looked at the long term effects of CAD trauma, the authors reported that the incidence of accelerated spondylosis was not statistically significant -- somewhat surprising in light of other reports. Their data was listed in tabular form and compared to normative age-matched data. Although the figures failed to reach significance at the $p < 0.05$ level, the incidence of degeneration in CAD patients was seen four times more frequently than in uninjured people. The significance level was $p = 0.07$ which means

that the probability that these differences were due simply to chance was only seven percent. Unfortunately, this type of confusion is often engendered by the use of p values and is one of the reason we are moving away from them.

In part IV of this series we'll look at how we can interpret our data and how far we can reasonably extrapolate these results.

References

1. Hassard TH: Understanding Biostatistics. St. Louis, Mosby Year Book, p 187, 1991.
2. Nachemson AL: Newest knowledge of low back pain: a critical look. Clin Orthop Rel Res 279:8-20, 1992.
3. Nachemson AL: Low back pain: are orthopaedic surgeons missing the boat? Acta Orthop Scand 64(1):1-2, 1993.
4. Rowland LP: Surgical treatment of cervical spondylotic myelopathy: time for a controlled trial. Neurology 42:5-13, 1992.
5. Riegelman RK: Studying a Study and Testing a Test: How to Read the Medical Literature. Boston, Little, Brown and Co., 1981.

Arthur Croft, DC, MS, FACO
Coronado, California

Click [here](#) for more information about Arthur Croft, DC, MS, MPH, FACO.



Page printed from:

http://www.chiroweb.com/mpacms/dc/article.php?id=42622&no_paginate=true&p_friendly=true&no_b=true