CEIPE THE CONSORTIUM OF EVIDENCE-INFORMED PRACTICE EDUCATORS

The Savvy Practitioner

A bulletin for practitioners and teachers of evidence-based practice.

The topic of this Savvy Practitioner was suggested by Dr.Hunnisett from McTimoney Chiropractic College. Do you have any requests for a specific topic? E-mail rlefebvre@uws.edu

Target audience this issue:

- ***Classroom faculty***
- ***Clinicians***
- ***Teaching material for EIP core instructors***

Ron LeFebvre DC
Executive Director CEIPE
University of Western States

Editorial suggestions by
John Hart DC, MHSc
Assistant Director of Research
Sherman College of Chiropractic

7/18/16
Issue 5

P Values and Statistical Significance

When looking at the results of a study, a natural question is—*are the reported results likely due to random chance alone?*

A quick and simple item to look at is the *p value*. The p value tells you the *probability* that the results were due to chance. That is, whether the differences being measured (e.g., between two treatment groups) are likely to be real and not simply due to randomness.

Reading P values

Reading p values is very easy.

.10	means that there is a 10% probability the results were due to random chance.
.04	means that there is a 4% probability that the results were due to random chance.
.001	means that the chances are only 1/10 of 1% probability that the results were due to random chance!

In health care research, it is generally agreed that to rule out chance alone as the reason for a result, the **p-value should be 0.05 or smaller**.

Sometimes when studies have very few subjects, a p value may be reported that is *justly barely* statistically significant at the 0.05 level (e.g., 0.04). In such cases even if just a few subjects had different outcomes, this could shift the p value towards "insignificance." A borderline p value in a small study can be a red flag.

On the other hand, even in a small study, if the p value is *very* small (e.g., 0.008), then the results may not be quite as vulnerable to the chance shift in the outcomes of just a handful of patients. In this case, the much smaller p value could be somewhat more reassuring. Although studies with very few subjects (e.g., < 30-50) are *still* generally problematic for a number of reasons.

When P values are > 0.05

When outcomes do not meet the 0.05 threshold, the statistics fail to adequately support the hypothesis that a particular treatment really works. The negative results should not, however, be immediately construed as evidence *against* the treatment. The results simply fail to rise to a level of adequate statistical certainty.

Help your students understand how to use read and understand the results from research studies.

Want to know more? Want handouts for your students?

Consult the Educator's Exchange!

Click through the following webpages: EIP Resources > Reading Results and download more materials on statistical significance.

Lost your link to the Educator's Exchange?

Try http://bit.ly/CEIPE.

You will need your password and user name.

Don't have access to the Educator's Exchange?

To sign up for this closed website, just contact rlefebvre@uws.edu and you will be sent an invitation to set a username and password.

The explanation could either be that **a) the treatment *really* doesn't have any beneficial effect or superiority** or **b) that the study size was too small to detect its true effect.**

Not enough power?

When the p value is greater than 0.05, *it is very important to see if the researchers reported a power calculation in which they predicted the number of subjects they needed to have in the study*. If they did report this, then check whether at the end of the study there were still enough patients left in the pool to maintain the requisite statistical power.

Traditionally, sample sizes are calculated to attain .80 power (i.e., an 80% chance that the study can detect a true benefit of one treatment over a control providing a true benefit actually exists). If the results fail to achieve *statistical significance,* this may have been because the study was too small to be able to detect the benefit.

In studies lacking sufficient power, the findings are not really evidence *against* the intervention, but rather the study simply *lacks the power to inform us* one way or the other.

Close, but….

What happens when a p value is just *barely* above the traditionally acceptable 0.05 (e.g., 0.06)? What to do in this scenario is more controversial.

Sometimes authors will indicate that there was a "trend toward" statistical significance. In most cases, researchers, peer reviewed journals, policy makers, etc. will judge the results to be statistically non-significant. In some hard sciences like physics, acceptable p values much lower than 0.05 are the norm.

The individual practitioner, however, has more leeway and may decide to be an early adopter of an intervention that showed great clinical benefit in a study while just missing the 0.05 threshold. Of course, many other elements must be weighed when making this decision such as the quality of the study itself and its size as well as the risks, benefits, costs and alterative interventions that may be available.

Others would caution against this strategy, especially for small studies. Many interventions that look promising even with much better p values do not ultimately prove to be valid. Ultimately an *informed* practitioner must decide for him or herself.

What p values do NOT tell you….

Beware! Even if the outcomes being measured *do* have statistically significant p values, it does not necessarily mean that the results of the study are *valid or that their benefits are large enough to be clinically significant* (i.e., large enough to really matter to a patient or a provider).

If outcomes are not due to chance alone, they still may not be a true reflection of the actual performance of a treatment. If the research design or its execution was flawed and bias creeps into the study, then the results may not at all be valid despite an acceptable p value.