

3. SCIENTIFIC PSYCHOLOGY AND WHAT TO DO WITH NON-STATISTICALLY SIGNIFICANT RESULTS

- 3.1. Non-Statistically Significant Results
- 3.2. Interpretive Bias
- 3.3. Footnotes
- 3.4. References

3.1. NON-STATISTICALLY SIGNIFICANT RESULTS

Scientific psychology is focused around the statistical testing of hypotheses. Finding statistically significant results is the ultimate achievement as well as being satisfying for the researcher. But what to do with results that are not statistically significant? It would be easy to conclude that such findings tell us nothing of use.

There are different types of non-statistically significant (NSS) results:

i) Findings that are as predicted but do not reach the level of statistical significance.

Generally in scientific psychology, a level of statistical significance of $p=0.05$ (5%) is accepted. This is as much convention as any absolute fact (1) (Watson 2004).

Put simply, and some will disagree with this description (2)(3), a statistical significance of 5% accepts that the data produced could be due to chance in five out of one hundred times. A study that finds their results to be $p=0.06$ (six in one hundred), for example, is not statistical significant.

"Just missing" statistical significance could be due to poor statistical testing, or the sample size, for example, and "really" the findings are statistically significant. It can be an example of a type II error (Neyman and Pearson 1933), particularly if the significance level chosen is too strict (4).

ii) Findings that are NSS and this tells us what could be statistically significant.

If researchers are attempting to isolate the causes of a behaviour, and there are many, knowing that one cause is NSS crosses that off the list and the next one can be tested.

iii) The results are NSS because there is no

difference between the groups.

This is the traditional meaning of NSS - the results occurred by chance. For example, in a clinical trial comparing a new psychotropic drug to a placebo pill, there is NSS difference found between them because the drug is no more effective than the placebo in reducing the disorder being studied. That is not to say that the drug has no effect, but that the effect is not statistically significantly greater than the placebo's effect.

In such clinical trials, there is a tendency to not publish such findings (Sterne and Davey Smith 2001). But the NSS results are as important as the statistically significant ones, or else a bias picture of the drug's effectiveness will be presented.

iv) Results that are NSS because they are completely opposite to the prediction.

For example, the one-tailed research hypothesis predicts group A to do better on a measure of behaviour, but it is found that Group B does better. The results being NSS here are very different to (i) above, for instance.

3.2. INTERPRETIVE BIAS

Hewitt et al (2008) are concerned about how the NSS results are interpreted. They highlight the risk of interpretive bias - overemphasis or underemphasis of results. Of course, interpretive bias is also relevant with statistically significant results.

For Hewitt et al the risk is greater with NSS results because the "observed difference may be real and the study is underpowered or the observed difference may occur simply by chance" (p23). In the case of randomised clinical trials, researchers are often "rarely neutral" because of their investment of "intellectual capital in developing the treatment under evaluation".

Hewitt et al looked at seven randomised clinical trials published recently in the "British Medical Journal" with NSS differences, and whether the authors still recommended the intervention being studied.

Taking one example, Henderson et al (2007). This was a comparison of "normal sex education" and a more detailed programme (SHARE) among 13-15 year-olds in twenty-five schools in Scotland. The outcome measure was a reduction in the NHS pregnancy termination rate.

The results were NSS because the intervention group

(SHARE programme) actual had an increase in terminations compared to the control group. The authors did not recommend withdrawal of the SHARE programme. They saw other benefits to the programme in areas like willingness to discuss condoms with partners, and intentions to resist unwanted sexual activities.

There may have been secondary benefits to the SHARE programme but it failed (ie NSS) on the main measure. Scientific psychology emphasises that it is that - scientific - and so must obey the rules of science. If findings are NSS, they are not "true". Such a belief is too simplistic, but if the aim is for psychology to be a science, the world has to be viewed like that. Those that live by the sword also die by it to use a common phrase.

That is why there are many benefits to psychology not being a science, and to making full use of qualitative methods. The psychology of teen pregnancy, terminations, and sex education are better understood by searching for the meaning that these hold for the individual rather than by numbers. But that is another story.

3.3. FOOTNOTES

1. Fisher (1950) is credited as the founder of the idea: "We shall not often be astray if we draw a conventional line at 0.05" (quoted in Sterne and Davey Smith 2001).
2. Coolican (2000) is concerned with the precision of explanations and terms used in teaching statistical testing.
3. Another common way to describe statistical significance is in terms of the probability that the null hypothesis is true.
4. Statistical significance is not the same as something being true. The results can be statistically significant but still due to chance because of the small probability of chance accepted. Likewise, NSS results could be true because there is always a small possibility that chance is not responsible for the findings.

3.4. REFERENCES

Coolican, H (2000) Hypothetically speaking
Psychology Teaching 8, 1, 3-9

Fisher, R.A (1950) Statistical Methods for Research Workers London: Oliver & Boyd

Henderson, M et al (2007) Impact of a theoretically based sex education programme (SHARE) delivered by teachers on NHS registered conceptions and terminations: Final results of cluster randomised trial British Medical Journal 334, 133-138

Hewitt, C; Mitchell, N & Torgerson, D (2008) Heed the data when results are not significant British Medical Journal 336, 23-25

Neyman, J & Pearson, E (1933) On the problem of the most efficient tests of statistical hypotheses Philosophical Transactions of the Royal Society A 231, 289-337

Sterne, J & Davey Smith, G (2001) Sifting the evidence - What's wrong with significance tests? British Medical Journal 322, 226-231

Whatson, T (2004) Beyond Reasonable Doubt: A Guide to Designing Experiments in the Behavioural Sciences Milton Keynes: Open University