Extracting meaning from research – using confidence intervals

Gissane, C.

School of Sport, Health and Applied Science, St Mary's University College, Twickenham, Middlesex, TW1 4SX, UK.

gissanec@smuc.ac.uk

Whenever disseminating research findings researchers should consider their audience, the reader.[1] To effectively transfer information to clinical practice, findings should be presented in an easy to understand format[2] that characterises the clinical importance of the findings.[3] In the first of the series of statistics bites the limitations of the 'P' value for decision making were outlined.[4] The 'P' value provides information as to whether or not the observed test result was due to chance. The role of chance does need to be evaluated, but relying on a P value alone is an over reduction of the data.[5] No matter how precisely the P value is reported it will not give an indication of the clinical importance of the results.[6] So, clinically important results may not be statistically significant, and significant results may not be clinically important. Study results must be presented in a way that clinicians can put them to use, estimation and the confidence interval (CI).

In reality, any study attempts to estimate what would be expected in a defined population of interest if they were all studied.[7] Estimation is the process of providing a numerical value for a population parameter.[8] It contrasts with hypothesis testing by aiming to quantify the effect of interest as an estimate that is clinically relevant.[5] In its simplest form, an estimate can be a single value or a 'point estimate', such as a mean or a percentage, but this gives no indication of how the value might vary in the population.[9] Around the point estimate, a CI gives a range of values, that are likely to contain the true population value at a given level of confidence, usually 95%.[2]

To illustrate the CI, the data presented in the first article[4] of this series will be used, it is re-shown in table1. The treatment group has undergone a new therapy for low back pain, and the visual analog scale (VAS) pain scores are reported. From the table it can be seen that in the standard care group VAS score changed by 0.1(0.3) and by 2.4(0.5) in the treatment group. When the null hypothesis of no difference between the groups' change scores was tested, it gave the results result t_{18} = -12.01, P = 0.001. There was a mean difference between the groups of 2.3 in favour of the treatment group. The 95% CI was calculated as 1.9 to 2.7.

So what is known? Firstly the difference is significant (P =0.001), with the information coming from the P value. Secondly, the best estimate of the difference in the two groups is the mean of 2.3. However, this will vary between 1.9 and 2.7, if the study was repeated.

The reader now has to decide if this is applicable to his or her practice. Firstly, is the observed difference clinically important? Is the observed difference noteworthy in this situation? Secondly, using the information from the 95% CI, would you as a practitioner be happy with the possibility that in the longer term that the score would vary between 1.9 and 2.7? Third, is this a good estimate, or is it too inconsistent for you? I would suggest that as the width of the 95% CI is 0.8, it is quite good. It is narrow and represents a good estimate of the population mean. Lastly, as the 95% CI does not cross the null value, in this case 0, the difference is significant, P < 0.05.

Table 1. Visual analog scale pain scores for standard care and treatment groups, pre and post.

	Standard Care			Treatment		
	Pre	Post	Change	Pre	Post	Change
1	4	4	0	4	1	3
2	4	4	0	6	3	3
3	5	5	0	5	3	2
4	3	3	0	3	1	2
5	6	5	1	6	3	3

6	4	4	*0*	4	2	2
7	2	2	*0*	3	1	2
8	3	3	*0*	3	1	2
9	6	6	*0*	4	1	3
10	2	2	*0*	3	1	2
Mean	3.9	3.8	*0.1*	4.1	1.7	*2.4*
SD	1.4	1.3	*0.3*	1.2	0.9	*0.5*

It is difficult to think of a situation where a CI is not preferable to a P value.[6] They convey lots of useful information for clinicians. A main function of a CI is to convey the (im)precision or uncertainty of the observed statistics.[3][5] At a given level of confidence, a narrower CI will provide a better estimate, with less variability than a wider CI.[6] The width of a CI is determined by three factors:-

1. The level of confidence chosen, a higher level of confidence will result in a wider interval;[10]

2. The inherent variability of want is being measured, less variability will lead to a more precise estimate;[10]

3. The sample size:[10] The variability is affected by the square root of the sample size.[5] Put simply, larger sample provides more information than smaller samples.

Another advantage for interpreting CIs is that are reported in the original units of measurement,[5] the clinician of doesn't have to decipher a P value. It is also possible for readers to calculate them if they are not provided.[5]

The International Committee for Medical Journal Editors (ICMJE) instructions for authors[11] requests CIs where appropriate, stating *"When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals). Avoid relying solely on statistical hypothesis testing, such as P values, which fail to convey important information about effect size."* In addition several authors have requested greater use of CIs[5][12][13] and a move toward estimation

science.[12] But, it has also been suggested that P values are reported alongside[6] as continuous numerical value[12] (e.g. P= 0.03). However, when reporting CIs, report the CI for the difference.[5] If individual group CIs are reported, their interpretation becomes much more difficult.[5] [14]

A CI is applicable whenever an inference to a population is being made. Their use moves decision making away from the yes or no of the P value. By presenting not a single value, but a range of values that estimate the population value, the reader is able to view them into context. Embracing evidence-based practice means integrating best research evidence with clinical expertise and patient values.[2] This puts the onus on clinicians to find and interpret research evidence. Similarly, researchers must present their findings in a way that they can be put to clinical use.[2] It has been suggested that clinicians are sometimes forced to make clinical decisions based on incomplete evidence.[15] To improve clinical decision making, researchers need to give information about the magnitude and certainty of the results, along with some indication of what is likely to happen in the population under investigation.[3] The time has come for researchers to present CIs and for readers to expect them.

References

1. Cumming G. *Understanding the new statistics: effects sizes, confidence intervals and meta-analysis.* London: Routledge, 2012.
2. Stratford PW. The added value of confidence intervals. *Physical Therapy* 2010;90:333-5.
3. Cordova ML. Giving clinicians more to work with: let's incorporate confidence intervals into our data. *Journal of Athletic Training* 2007;42:445.
4. Gissane C. The P value, do you know what it means? *Physiotherapy Practice and Research* 2012;33(2):105-6.
5. Altman DG. Why we need confidence intervals. *World Journal of Surgery* 2005;29:554-6.
6. Porcher R. reporting results of orthopaedic research. *Clin Orthop Relat Res* 2009;467(10):2736-7.
7. Lang TA, Secic M. *How to report statistics in medicine.* Phiadelphia, PA: American College of Physicians, 1997.
8. Everitt BS. *Dictionary of Statistics in Medical Sciences.* Cambridge: Cambridge University Press, 1995.
9. Bland M, Peacock J. *Statistical questions in evidence based medicine.* Oxford: Oxford University Press, 2000.
10. Gardner MJ, Altman DG. Estimating with confidence. In: Altman DG, Machin D, Bryant TN, Gardner MJ, editors. *Statistics with confidence.* London: BMJ Books, 2000:3-5.

11. ICMJE. Uniform requirements for manuscripts submitted to biomedical journals: Preparing a manuscript for submission to a biomedical journal 2009;Available from: http://www.icmje.org/manuscript_1prepare.html (Accessed 6th September 2012).

12. Marshall SW. Testing with confidence: use (and misuse) of confidence intervals in biomedical research. *Journal of Science and Medicine in Sport* 2004;7:135-37.

13. Wolfe R, Cumming G. Communicating uncertainty in research findings: confidence intervals. *Journal of Science and Medicine in Sport* 2004;7:138-42.

14. Wolfe R, Hanley J. If we're so different, why do we keep overlapping? when 1 plus doesn't make 2. *CMAJ* 2002;106:65-6.

15. Doody C. Evidence based practice requires sound clinical reasoning. *Physiotherapy Ireland* 2011;32:4-5.