Brazilian Journal of Physical Therapy

https://www.journals.elsevier.com/brazilian-journal-of-physical-therapy

EDITORIAL

Showing confidence (intervals)

Introduction

Confidence intervals take centre stage in two articles in this edition of the journal. Freire et al.[1] report on how often physical therapy trials use confidence intervals to express between-group differences, and Hespanhol et al.[2] present a Masterclass on the mathematical underpinning and interpretation. Two important points are contained in these articles: (1) confidence intervals provide critical information in research reports, and (2) interpretation of confidence intervals is a key component of evidence-based physical therapy practice. This editorial aims to provide a brief 'user's guide' for clinicians looking to incorporate research findings into practice.

Why use confidence intervals?

Health research has traditionally used the *p*-value to characterise differences between groups as statistically significant or not. Despite recognition of theoretical and statistical problems with *p*-values decades ago, the clinical research world has been slow to move towards better methods. The main problems for a clinician trying to use research reported with *p*-values to inform their practice are as follows:

- Accurate interpretation of a *p*-value is complicated and not intuitive;
- A *p*-value by itself gives no information about the size of a treatment effect;
- Statistical significance of a *p*-value is sensitive to sample size, bigger studies have smaller *p*-values for the same sized effect.

Effect estimates with confidence intervals provide a different way of describing the difference between groups. A paper that includes confidence intervals has important benefits for clinicians wishing to use research to inform their practice. Effect size estimates with confidence intervals provide two critical pieces of information:

1. An estimate of the size of the mean effect of treatment;
2. An indication of how precise the effect estimate is.

This information enables a clinician to make judgements about whether they should incorporate the results of a study into practice. There is gathering consensus among researchers that between-group differences should be accompanied by confidence intervals rather than *p*-values, as Freire et al.[1] write; ''*a statistically significant finding should not be interpreted on its own to influence clinical practice*''. Their study also shows that there is a shift underway, as increasing numbers of RCTs on the PEDro database report effects with confidence intervals.

How to use confidence intervals

To understand why a study provides an estimate (rather than a definitive answer) requires recognition of the difference between a sample (people in the study) and a population (people to whom the study results are applied). Any single study cannot give an exact prediction of what will happen to all the people in the population; that is why we call the mean between-group difference an 'estimate'. The confidence interval around this estimate provides a way of showing the range of values within which the true effect probably lies. When confidence intervals are narrow, we can be relatively certain about how effective a treatment is, but when they are wide, the opposite is true. Notwithstanding the nuances surrounding different analytical approaches detailed in Hespanhol et al.,[2] some rules of thumb can be applied to using confidence intervals.

Worthwhile effects

To apply research results to practice requires attention to the concept of 'worthwhile difference'. Defining and quantifying worthwhile difference is a whole subject of itself, but at its core, the idea is simple. A worthwhile difference is the minimum amount of benefit a person would need
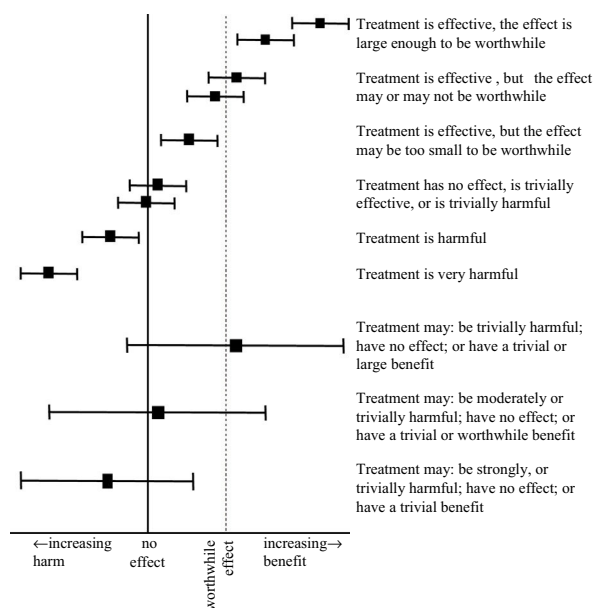
Figure 1 Interpreting confidence intervals around between-group differences. In the top half of the figure, the confidence intervals are narrow. Narrow confidence intervals are informative; they enable definite statements about the estimate of the treatment effect. In the bottom half of the figure, the confidence intervals are wide. Wide confidence intervals are less informative; they do not allow definite statements about the estimate of the treatment effect.

to experience, to make receiving the treatment beneficial. All treatments involve time, effort, money, attention and risk, and therefore there has to be a big enough improvement to make the investment worth it for the patient. For example, Areeudomwong and Buttagat[3] tested a specific neuromuscular facilitation exercise program that required close supervision and feedback from a physical therapist versus a set of simple trunk exercises for people with chronic back pain. They found an improvement of 1.2 points on a 0–10 pain scale. The question is whether the extra effort, specialised supervision (and potentially cost) is worth it for an effect of this size.

Researchers have tried to define how big an effect must be for it to be worthwhile for some outcome measures. Doing this is a quite difficult task for several reasons, so there is no universal agreement on the size of a worthwhile effect. In a clinical situation, it might be possible to discuss the estimated effect directly with a patient; this is a way to incorporate patient preferences into treatment decisions.

Upper and lower bounds

Interpretation of effects involves looking at where the estimate of effect and the upper and lower bounds of the confidence interval sit in relation to the meaningful effect (Fig. 1). The confidence intervals tells the reader that the true effect of treatment likely lies somewhere within this range, probably nearer the centre. This means looking at whether the upper and lower bounds of the confidence interval cross over the line of meaningful difference, which tells us something about how likely it is a treatment will be effective, effective but not worthwhile (trivial effective), ineffective, or harmful (less effective than control).

Acknowledgements

References

1. Freire APCF, Elkins MR, Ramos MC, Moseley AM. Use of 95% confidence intervals in the reporting of between-group differences in randomized controlled trials: analysis of a representative sample of 200 physical therapy trials. *Braz J Phys Ther*. 2019;23:302–310.

2. Hespanhol L, Vallio CS, Saragiotto BT, Costa LCM. Understanding and interpreting confidence and credible intervals around effect estimates. *Braz J Phys Ther*. 2019;23:290–301.

3. Areeudomwong P, Buttagat V. Proprioceptive neuromuscular facilitation training improves pain-related and balance outcomes in working-age patients with chronic low back pain: a randomized controlled trial. *Braz J Phys Ther*. 2018, http://dx.doi.org/10.1016/j.bjpt.2018.10.005 [in press].

Steven J. Kamper[a,b,*]
ª *University of Sydney, School of Public Health, Institute for Musculoskeletal Health, Australia*
ᵇ *Centre for Pain, Health and Lifestyle, Australia*

* Correspondence to: Institute for Musculoskeletal Health, PO Box M179, Missenden Rd, Camperdown, NSW 2050, Australia.
E-mail address: Steven.kamper@sydney.edu.au