Interpreting a meta-analysis: what should we focus on?

Having read the recent systematic review by Booth et al., Noelle Moreau argues between-group effect sizes and forest plots should be reported in future systematic reviews and meta-analyses.

Editor,

Interpreting and reaching conclusions from a systematic literature review and meta-analysis can be challenging. A meta-analysis combines results from randomized controlled trials (RCTs) and provides summary statistics, or effect sizes, for individual studies, as well as a summary effect size across similar studies.[1] RCTs compare an experimental intervention to an alternate 'control' intervention (placebo, no intervention, or standard of care). RCTs thus allow us to determine whether there is a cause-and-effect relationship between the independent and dependent variables. Most threats to internal validity can be mitigated through randomization of subjects to groups and by having a control or comparison group that is similar in baseline characteristics to the experimental group.[2]

If we over-interpret the data and report *within-group* effect sizes alone, we ignore the basic premise for an RCT. This can result in threats to internal validity, of which the most notable are history and maturation effects, especially in studies involving children. Simply put, if we only report pre- to post-intervention changes in the experimental group, we have no way of knowing if the outcome of interest improved simply due to maturation or passage of time, or due to the effect of our intervention. However, if the experimental group change is greater than the change observed in the control group and is statistically significant, we can be confident that the outcome was likely due to the experimental intervention and not simply to maturation or natural history. This is why *between-groups* effect sizes, typically reported and graphed in forest plots, are so important.

Figure 3 in Booth et al.[3] displays a typical forest plot comparing the outcomes of gait training to control/standard of care. Contrast this to Figure 2, which displays *within-group* effect sizes. Interpreting these within-group effect sizes without comparison to the control intervention is subject to the threats to internal validity described earlier. For example, Booth et al. report on page 9 that improvements in gait speed after gait training were maintained up to 6 months post-intervention, and cite Aviram et al.[4]. However, readers may be surprised to know that gait speed improved in the control group to a greater extent from pre- to post- compared to the treadmill training (TT) intervention, and that these improvements were also maintained at 6 months. Examination of Figure 2 alone (illustrating *within-group* effect sizes) does not communicate this information. Visually, the effect size (standardized mean difference indicated by a diamond) for 'PT pooled' (control) appears very similar to the effect size (diamond) for 'TT pooled' for Aviram et al.[4] Only when examining the *between-group* effect sizes and forest plot in Figure 3 do we see clearly that for Aviram et al.,[4] the effect size (or standard mean difference) of .40 favors the control group in that study for the outcome of gait speed.

Booth et al.[3] reported significant improvements in the GMFM after gait training and that positive effects in the GMFM were retained up to 6 months post-intervention, again citing Aviram et al.[4] as the sole source. This interpretation should be greatly tempered when we factor in that the improvements in the GMFM were significantly greater in the control group as compared to the treadmill training group, and that this difference was maintained at 6 months in favor of the control group!

These examples illustrate that the interpretation of *within-group* changes alone is a slippery slope that can often be misleading. For this reason, I believe that future systematic reviews and meta-analyses should focus on the standard *between-group* effect sizes and forest plots, in order to take advantage of the design advantages of RCTs and to avoid the threats to internal validity that occur with over-interpretation of within-group effect sizes.

Noelle G Moreau, PhD, PT
Associate Professor, Department of Physical Therapy, Louisiana State University Health Sciences Center – New Orleans

References

1. Borenstein M. Introduction to Meta-Analysis. Chichester, U.K.: John Wiley & Sons, 2009.
2. Portney LG, Watkins MP. Validity in experimental design. In: Foundations of Clinical Research: Applications to Practice, 3rd edition. Upper Saddle River, N.J.: Pearson/Prentice Hall, 2009.
3. Booth ATC, Buizer AI, Meyns P, Lansink ILB, Steenbrink F, van der Krogt MM. The effect of functional gait training in children and young adults with cerebral palsy: a systematic review and meta-analysis. Dev Med Child Neurol 2018 Mar 7. doi: 10.1111/dmcn.13708. [Epub ahead of print].
4. Aviram R, Harries N, Namourah I, Amro A, Bar-Haim S. Effects of a group circuit progressive resistance training program compared with a treadmill training program for adolescents with cerebral palsy. *Dev Neurorehabil* 2017; **20**: 347–54.