

Commentary

Between-Groups Study Designs Demand Between-Groups Analyses: A Response to Hernandez-Reif, Shor-Posner, Baez, Soto, Mendoza, Castillo, Quintero, Perez, and Zhang

Christopher A. Moyer

University of Wisconsin-Stout

The most recent massage therapy (MT) study by Hernandez-Reif *et al.* displays flaws persistent in this area of research that are attributable to MT researchers' frequent mistake of using within-group analyses of dependent variables in studies that are purported to be randomized control trials. This practise violates the logic of using randomization to create treatment and control groups, and thereby fails to control for the validity threats of spontaneous remission, placebo effects, and statistical regression. The result is that a clear understanding of what MT can and cannot do is seriously hampered.

Keywords: Massage – cortisol – methodology – randomized – control group

The appearance of a new randomized control trial of massage therapy (MT) should always be a positive event. This is true whether or not MT was observed to perform better than a control, because any randomized control trial that is well conducted is an important contribution to the progressive and systematic accumulation of knowledge that defines science. However, too often in MT research the appearance of a new randomized control trial is a mixed blessing, or even a missed opportunity. This is because randomized control trials of MT frequently have major flaws (1,2), including the failure to accurately review previous findings and the failure to follow the logic of their own research design. I was reacquainted with these persistent flaws when I read Hernandez-Reif *et al.*'s (3) recent report in this journal on the effects of MT on Dominican children with HIV. Because I believe it is critically important for the state of MT research to improve, I highlight these flaws.

Failure to Accurately Review Previous Findings

In reviewing previous research, Hernandez-Reif *et al.* contend unequivocally that MT 'has been shown to

enhance immune functions... [of] children' and speculate that this follows from MT's effect on the stress hormone cortisol. This is misleading, as extant data do not support these effects. Quantitative reviews of MT randomized control trials do not yield statistically significant effects on cortisol levels of adults (4) [effect size (ES) = 0.14, 95% confidence interval (CI) = -0.10, 0.38] or children (5) (ES = 0.28, 95% CI = -0.27, 0.84), nor on childrens' immune functions (5) (ES = 0.06, 95% CI = -0.52, 0.63).

Presumably, Hernandez-Reif *et al.* have scientifically grounded reasons to disagree with the results of those quantitative reviews. In that case, those reasons should be illustrated or at least mentioned in their section asserting MT effects on cortisol and immune functions. What should not happen is for MT effects that are contentious to be presented as well established. Even a brief review of the results on which a current study is based should accurately represent the consensus, or lack thereof, among researchers.

Hernandez-Reif *et al.*'s support for cortisol and immune function effects, and my own position that these effects are unestablished and possibly nonexistent, are primarily based on the same set of studies. Why do we reach opposite conclusions from the same set of studies? I am convinced that the cause of this discrepancy is what I discuss next; specifically, that randomized

For reprints and all correspondence: Christopher A. Moyer, University of Wisconsin-Stout, Tel: 715-232-1621; Fax: 715-232-5303; E-mail: christopher.a.moyer@gmail.com

© 2007 The Author(s).

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

control trials of MT are frequently analyzed and reported *as if they were uncontrolled within-group studies*, which consequently leads to misinterpretation of their results. In effect, they are randomized control trials in name only, and fail to properly utilize the well-established logic of randomization and experimental control.

Failure to Follow the Logic of Randomization and Control

In discussing their results, the study authors state they have completed 'the first randomized control trial to examine massage therapy for enhancing development and decreasing maladaptive behaviors in young Dominican children infected with HIV,' but this is not true, for the study is not a randomized control trial if it does not make between-groups comparisons of the dependent variables. The entire purpose of employing randomization to create treatment and control groups, as opposed to using a simpler within-group design, is so that one can make *between-groups* comparisons that control for the validity threats of spontaneous remission, placebo effects, and statistical regression.

Having performed analyses inconsistent with the study design, and that introduce rather than control these threats, the authors reach conclusions that are difficult to justify. To conclude that MT improved the behavior of the older sample of children, that it provided a 'marginally significant' increase in IQ, or that it 'appears to be a viable therapy for promoting greater daily functioning and communication' based on within-group analyses is misleading, because any or all of those pre-post effects could be observed as a result of the aforementioned threats even if MT was wholly ineffective. Further, those threats cannot be controlled by the authors' occasional use of what could be called 'side-by-side' within-group comparisons, in which they test each group separately for its own pre-post effect, and then give the impression that MT worked if it shows a statistically significant pre-post effect and the control group does not. This approach is not at all equivalent

to making the required between-groups comparison, and the result is misleading, as it is easily possible in that situation (especially with small samples) to have only one of the two within-group comparisons be statistically significant when there is actually no difference between the groups.

Possibly the authors' decision to exclude between-groups analyses of dependent variables was motivated by concern that the tests would be underpowered and therefore not attain statistical significance. However, even if this was the case, the small size and exploratory nature of the study cannot be used to justify the decision to use within-group analyses. If problems pertaining to the statistical power of between-groups analyses were anticipated, there are defensible alternatives. These include choosing a more liberal value for alpha (e.g. using $P < 0.10$ or $P < 0.20$, rather than $P < 0.05$), placing greater emphasis on effect sizes and their confidence intervals than on probability values, and examination of clinical significance (6). Simply stated, between-groups study designs such as randomized control trials logically demand between-groups analyses. Until this is reflected in MT research, our knowledge of what it does and does not do will continue to be seriously hampered.

References

1. Ernst E. Massage therapy for low back pain: a systematic review. *J Pain Symptom Manage* 1999;17 (1): 65-9.
2. Furlan A, Brosseau L, Imamura M, Irvin E. Massage for low-back pain: a systematic review within the framework of the Cochrane collaboration back review group. *Spine* 2002;27 (1): 1896-1910.
3. Hernandez-Reif M, Shor-Posner G, Baez J, Soto S, Mendoza R, Castillo R, et al. Dominican children with HIV not receiving antiretrovirals: massage therapy influences their behavior and development. *Evid Based Complement Alternat Med* (in press).
4. Moyer CA, Rounds J, Hannum JW. A meta-analysis of massage therapy research. *Psychol Bull* 2004;130 (1): 3-18.
5. Beider S, Moyer CA. Randomized controlled trials of pediatric massage: a review. *Evid Based Complement Alternat Med* 2007;4 (1): 23-34.
6. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psych* 1991;59 (1): 12-19.

Received June 15, 2007; accepted September 19, 2007