Article:
Morley, S (2014) Count up to 2 but not beyond. Pain. ISSN 0304-3959

https://doi.org/10.1016/j.pain.2014.03.020

Editorial Commentary

Accepted for Publication in Pain 2014

'Count up to two but not beyond'[1]

Stephen Morley

University of Leeds

Correspondence to

Stephen Morley

T: +44 113 343 2733

E: s.j.morley@leeds.ac.uk

Address

Charles Thackrah Building, University of Leeds, 101 Clarendon Road, Leeds, LS2 9LJ, UK

Psychologists have a strong preference for measuring human attributes (emotions, behavioral repertoires, cognitive appraisals) as dimensional continuities, routinely testing for differences with a Fisherian null hypothesis procedure to detect mean difference. The importance of that difference has often been of secondary concern. There have been plenty of dissenting voices [1; 9; 20] but old habits die hard. One consequence of this is that the evaluation of treatment effectiveness, at least since Glass's pioneering meta-analysis [14], has been expressed as an effect size: the magnitude of differences between two groups expressed as a ratio of their variance. There are two points to note. First, the effect size is a statistical measure devoid of any additional meaning; in and of itself it has no external referent. Second, it is difficult to express the benefits of a treatment in a way that is readily understandable to patients or even to other stakeholders. An effect size of 0.8, an omnibus value for psychological treatments [19], is the 79th percentile of the control group. An effect size of around 0.2, a current estimate for psychological treatments of pain corresponds to the 58th percentile [10]. This is difficult to communicate as it doesn't correspond to the patient's reasonable enquiry 'how many people get better'. In this case counting to two ('better' vs. 'not better') might be a strategy to be preferred to divining subtle between group differences.

Few trials of psychological interventions count to two. Where they have done so, the tendency has been to dichotomize a continuous measure [16]. This problem is not unique to psychological measures. There is a copious literature on dichotomizing and categorizing continuous measures of pain. Valid cut scores for mild, moderate, and severe pain are established by testing the association of pain scores with indices of functional performance [3; 5; 12]. In this issue of PAIN Sil and colleagues [13] report a secondary analysis of trial data where the original hypothesis was that a CBT intervention would 'be significantly more effective' (p.298) than education in reducing the primary outcome of functional disability in children with juvenile fibromyalgia i.e., an hypothesis of difference in means [8]. The immediate post treatment

effect gave an impressive uncorrected effect size = 0.78 (data in Table 2 [8]), which is good in the context of similar trials [4].

Sil et al take a two stage approach to evaluating clinical significant by following Jacobson's methodology [6]. Step one establishes the magnitude of change required for the change to be beyond that attributable to measurement error. This smallest detectable value is a statistical criterion for measurement precision not a clinical criterion. The setting of clinical criteria is more problematic and varied. Jacobson's method uses the statistical properties of the measure to set three possible cut points. For two cut points data from a non-clinical reference sample is required. The third cut point relies solely on the distribution of the clinical group. Each cut point categorizes the individual in reference to their relative position in the group. The cut points are determined by statistical criteria intrinsic to the measure. The validity of the interpretation depends on the likely assignment to group membership. As such it may tell us little about other parameters likely to be of clinical or social importance e.g. probability of relapse.

Sil et al take a slightly different approach to defining clinical significance for the Functional Disability Inventory (FDI). They use cut points established in a prior analysis of data from 1300 children with a variety of pain disorders [7]. That study set cut points at the upper and lower quartiles of the distribution to give three groups (no/minimal, moderate and severe disability) validating the classification by correlating FDI scores with pain intensity. This is somewhat ironic given the history of of validating pain scores by reference to function (*supra*). Arguably validation of a measure for clinical significance would be better established using operations of convergent validity i.e. other measures of function.) Kashikar-Zuck et al [7] analysis of the FDI showed that mean disability increased with age and that there were slight differences between clinical groups in FDI score: information not used in Sil et al's analysis.

There are many advantages in using the distributional properties of scales in outcome trials. Our use of them would be aided by the development of high quality norms, including

parameters such as age, gender, diagnostic group (including non-clinical), and country, which capture known variation in data [11; 17]. Larger samples, obtained by combining data sets with meta-analytic techniques will give more accurate estimates and better prediction for individuals. The work of IMMPACT [2; 15; 18] has focused attention on the need rationalize the plethora of measures used and this should facilitate the selection of measures for further development. Validation, especially with regard to establishing meaningful clinical cut points i.e. those that enable us to make good clinical decisions, is an ongoing requirement. We need to be able to count to two but we also need to capture the subtlety and variation in human behavior.

Disclosure

The author declares no conflicts of interest.

Footnote on the title

[1] The title is a quotation from George W Pickering's text (*High Blood Pressure*, 2nd edition, London:J&A Churchill, Ltd., 1968) and restated by the British epidemiologist Geoffrey Rose. See an imaginary, and enlightening, conversation between Rose and Socrates at: http://alertandoriented.com/geoffreyrosepart3/Rose - accessed 20 March 2014

References

[1] Cohen J. The earth is round (p < .05). Amer Psychol 1994;49(12):997-1003.

[2] Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, Kerns RD, Stucki G, Allen RR, Bellamy N, Carr DB, Chandler J, Cowan P, Dionne R, Galer BS, Hertz S, Jadad AR, Kramer LD, Manning DC, Martin S, McCormick CG, McDermott MP, McGrath P, Quessy S, Rappaport BA, Robbins W, Robinson JP, Rothman M, Royal MA, Simon L, Stauffer JW, Stein W, Tollett J, Wernicke J, Witter J. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113(1-2):9-19.

[3] Farrar JT, Young JP, Jr., LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001;94(2):149-158.

[4] Fisher E, Heathcote L, Palermo TM, de C Williams AC, Lau J, Eccleston C. Systematic Review and Meta-analysis: Psychological Therapies for Children with Chronic Pain. J Pediatr Psychol 2014. doi: 10.1093/jpepsy/jsu008.

[5] Hoffman DL, Sadosky A, Dukes EM, Alvir J. How do changes in pain severity levels correspond to changes in health status and function in patients with painful diabetic peripheral neuropathy? Pain 2010;149(2):194-201. doi: 10.1016/j.pain.2009.09.017.

[6] Jacobson NS, Roberts LJ, Berns SB, McGlinchey JB. Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. J Consult Clin Psychol 1999;67(3):300-307.

[7] Kashikar-Zuck S, Flowers SR, Claar RL, Guite JW, Logan DE, Lynch-Jordan AM, Palermo TM, Wilson AC. Clinical utility and validity of the Functional Disability Inventory among a multicenter sample of youth with chronic pain. Pain 2011;152(7):1600-1607. doi: 10.1016/j.pain.2011.02.050.

[8] Kashikar-Zuck S, Ting TV, Arnold LM, Bean J, Powers SW, Graham TB, Passo MH, Schikler KN, Hashkes PJ, Spalding S, Lynch-Jordan AM, Banez G, Richards MM, Lovell DJ. Cognitive behavioral therapy for the treatment of juvenile fibromyalgia: A multisite, single-blind, randomized, controlled clinical trial. Arthritis & Rheumatism 2012;64(1):297-305. doi: 10.1002/art.30644.

[9] Meehl PE. Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald and the slow progress of soft psychology. J Consult Clin Psychol 1978;46:806-834. doi: 10.1037/0022-006X.46.4.806.

[10] Morley S, Williams A, Eccleston C. Examining the evidence about psychological treatments for chronic pain: time for a paradigm shift? Pain 2013;154(10):1929-1931. doi: 10.1016/j.pain.2013.05.049.

[11] Roelofs J, van Breukelen G, Sluiter J, Frings-Dresen MH, Goossens M, Thibault P, Boersma K, Vlaeyen JW. Norming of the Tampa Scale for Kinesiophobia across pain diagnoses and various countries. Pain 2011;152(5):1090-1095. doi: 10.1016/j.pain.2011.01.028.

[12] Serlin RC, Mendoza TR, Nakamura Y, Edwards KR, Cleeland CS. When is cancer pain mild, moderate or severe? Grading pain severity by its interference with function. Pain 1995;61(2):277-284.

[13] Sil S, Arnold LM, Lynch-Jordan A, Ting TV, Peugh J, Cunningham N, Powers SW, Lovell DJ, Hashkes PJ, Passo M, Schikler KN, Kashikar-Zuck S. Identifying Treatment Responders and Predictors of Improvement after Cognitive-Behavioral Therapy for Juvenile Fibromyalgia. Pain. doi: 10.1016/j.pain.2014.03.005.

[14] Smith ML, Glass GV. Meta-analysis of psychotherapy outcome studies. Am Psychol 1977;32(9):752-760.

[15] Turk DC, Dworkin RH, Revicki D, Harding G, Burke LB, Cella D, Cleeland CS, Cowan P, Farrar JT, Hertz S, Max MB, Rappaport BA. Identifying important outcome domains for

chronic pain clinical trials: An IMMPACT survey of people with pain. Pain 2008;137(2):276-285.

[16] Turner JA, Mancl L, Aaron LA. Short- and long-term efficacy of brief cognitive-behavioral therapy for patients with chronic temporomandibular disorder pain: a randomized, controlled trial. Pain 2006;121(3):181-194. doi: 10.1016/j.pain.2005.11.017.

[17] Van Breukelen GJ, Vlaeyen JW. Norming clinical questionnaires with multiple regression: the Pain Cognition List. Psychol Assess 2005;17(3):336-344. doi: 10.1037/1040-3590.17.3.336.

[18] von Baeyer CL, Spagrud LJ. Systematic review of observational (behavioral) measures of pain for children and adolescents aged 3 to 18 years. Pain 2007;127(1-2):140-150.

[19] Wampold BE. The great psychotherapy debate: Models, methods, and findings. Mahwah, NJ: Lawrence Erlbaum Associates, 2001.

[20] Ziliak ST, McCloskey DN. The Cult of Statistical Significance: How the Standard Error Costs us Jobs, Justice and Life. Ann Arbor: The University of Michigan Press, 2008.