This is an author produced version of *Count up to 2 but not beyond*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/78820/

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

http://dx.doi.org/10.1016/j.pain.2014.03.020
Editorial Commentary
Accepted for Publication in Pain 2014

‘Count up to two but not beyond’

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 reference. 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 to 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

References


