Reviewer's report

Title: Agreement between pre-post measures of change and transition ratings as well as then-tests

Version: 2 Date: 27 September 2012

Reviewer: Wilco Emons

Reviewer's report:

Review BMC Medical Research Methodology:
Agreement between pre-post measures of change and transition ratings as well as then-tests

Overall, I think the manuscript has improved considerably. Both the structure and the focus are now much clearer. Nevertheless, I still have a few comments left (see below), the most important ones concern the statistical analyses, but I think they can be tackled quite easily.

Major Compulsory Revisions

1. (p. 10) Description of the direct measures of change. It is still not completely clear to me how direct measures of change were obtained. On page10 the authors state that “For each item, the response format for the direct measures of change comprised five categories (1 - markedly better, 2 - slightly better, 3 – no change, 4 - slightly worse, 5 - markedly worse). We first calculated the mean of the single-item ratings that belong to one of the three outcome scales (sleep, physical functioning, somatization).” What does ‘for each item’ in the first sentence refer to? Did you use a single item for each domain (sleep, physical, and somatization), thus in total 3 items, or multiple items per domain (e.g., taken from the IRES, SF36 or SCL90R)? (As the second sentence seems to suggest), as the remark in the discussion “We focused on three self-reported outcome domains (sleep, physical functioning, and somatization) for which the three different approaches to measuring change of interest to us were based on scales with equal numbers of items and equivalent content” suggests. I think this part needs further clarification (thus, how many items were used, how did you select items, etc.).

2. (p. 11) Recall bias. I’m little concerned whether correlations between t0 and retrospective scores are appropriate measures of the presence or absence of recall bias. Correlations are invariant under linear transformations. If there is a consistent recall bias (say all patients are consistently somewhat more optimistic), the correlation will be high, but recall bias is present (i.e., there is mean shift; changes scores are biased downwards). This means that if there is a high correlation, respondents can recall their relative position to others quite well, but not necessarily their true status. But if I understood it correctly, t0 and retrospective outcomes are obtained with the same questionnaires, right? If that
is the case, you could examine absolute agreement (ICC), or mean shift (e.g.,
dependent samples t-test). This may also help to substantiate the conclusion that
“Recall bias did not appear to play a major role in this regard”, which indeed is
the case.

3. (p. 11) Analysis of present state effect. This is an interesting phenomenon. I
think the regression approach used here, which was proposed by Guyatt et al.,
seems a neat idea, although the interpretation of the coefficients as to whether a
present-state effect exist may not always be clear cut in real data analyses.
However, I was wondering why you didn’t regress transition scores on indirect
change and post test scores; that is, you look for association with posttest (X2)
controlling for “true change”. If present state effect is absent, than X2 has no
partial effect on the transition scores; and if present-state effect exists, X2 should
have a partial effect. In its extreme, if respondents only use present-state status
to estimate their change direct, the indirect change scores have no partial effect
on transition scores. In my view, these results are more straightforward than
comparing two beta weights. And I think that you don’t have to rely on the equal
variance assumption, which is quite restrictive, particularly when there are
differential treatment effects (some patients change more than others).

Discretionary Revisions

1. (p. 11) I would remove the sentence “We did not use the intra-class correlation
coefficients
because the different measures were not on the same scale”.

2. (p.16). Limitation; bias due to pretesting: this relates to one of my major
comments in the review of the first draft. I agree with the author(s) response and
corresponding explanation in the discussion.

Table 2. SCL90R: I’m not familiar with all the details of these instruments, but
you may add the direction of the scale (e.g., higher scores indicate more/less
psychosomatic complaints). If I look in Table 2, it seems as if direct ratings are
opposite to those for (quasi) indirect given the differences in signs. I assume the
explanation is the direction of the scale.

(p. 14) “Discussion: Before substituting the quasiindirect approach of measuring
change for the indirect approach across the board,” What do you mean?

Level of interest: An article whose findings are important to those with closely
related research interests

Quality of written English: Needs some language corrections before being
published

Statistical review: Yes, and I have assessed the statistics in my report.
Declaration of competing interests:

I declare that I have no competing interests.