Reviewer's report

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

Version: 1 Date: 23 February 2012

Reviewer: Felix Angst

Reviewer's report:

Positive criticism
This study compares 3 different methods of measuring change of health over time. This is an interesting issue and the validity of assessment of change is important.

Specific positive issues are:
1. Clear and informative presentation of the 3 concepts of measurement, especially Figure 1. However, the Introduction could be shortened.
2. Clear and complete data presentation, especially by Table 2. However, I do not think than it is important to know all socio-demographic data (e.g., marital status, smoking etc.) of Table 1 to answer the study’s aim.

Major compulsory revisions
1. The article needs careful English editing. Many sentences are hard or impossible to understand and many statements remain unclear. Many terms could be more accurate and specific. Some examples:

P. 11, Discussion, 1st sentence: “First of all, this study shows that we cannot expect a clear-cut pattern of differences between effect sizes of the three different measures of change.” What is a “clear cut pattern”? What was your hypothesis?

P. 11, Discussion, 2nd sentence: “Contrary to previous studies, ...”. Better: “In contrast to...” And: write in the past!

P. 11 ff. The discussion is, in some parts, confusing and unclear. Especially the first paragraph. This part should summarize what you have done in your study, the main results, and the most important interpretations.

P. 12: “Quasi-indirect measures of change are more efficient to obtain that indirect measures of change”. What means “efficient”; with respect to what?

P. 12: “The correlation coefficients were not in the realm of the high value reported by Middle et al. but in all three scales analyzed at the upper end of those values reported by Kohlmann and Raspe”. Meaning of the whole sentence is unclear.

P. 13: “To our knowledge there is only one study to analyse ....”. Possibly better:
“To our knowledge, there exists only one study that analyzed ...”

P. 1, Title: “Agreement” or “correlation” are possibly more precise/accurate than “correspondence”.

2. The major concern is about the metric measurement properties of the 3 assessments of change. From the methodological point of view, it is not possible to compare the changes assessed by the transition item and those assessed by the indirect and quasi-indirect measure by means of parametric Pearson correlations. Non-parametric Spearman rank correlation may be more appropriate or a stratified analysis within the strata of the response categories of the transition item. Possibly, the best solution would be the use of an item-response-theory technique. Two major reasons for that:

1. The metric properties of the transition item (direct measure of change) are completely different than those of the indirect and quasi-indirect measures. The transition item assesses the change of health by 5 levels (P. 9: markedly better, slightly better, same, slightly worse, markedly worse) whereas the indirect and quasi-indirect measures use (changes/differences of) multi-item scaled scores, i.e., have many possible levels.

2. The interval of change of (for example) “markedly better” to “slightly better” is not the same (in terms of metrical distance) as that of “slightly better” to “same”. However, the intervals are taken into analysis as equally large (delta=1), see p.9, 1st paragraph.

3. In this context of item 2; P. 9. “Effect size for the direct measure (transition rating) were (remark: should be “was” or “effect sizes”) calculated by dividing the mean change score by its standard deviation.” This means that you used the standard deviation (SD) of the change. That corresponds to a “standardized response mean (SRM)”, i.e. the Hedges’ g for 1 sample (see: Rosenthal R. Parametric measures of effect size. In: Cooper H, Hedges LV (Eds.). The Handbook of research synthesis. Russell Sage Foundation, New York 1994;16:pp.231-44). This stays in contrast to the effects sizes (ES) of the indirect and quasi-indirect measures which are the score difference divided by the SD of the scores at baseline. To be comparable, one should either use ES or SRM for all 3 measures.

4. Agreement of the 3 different concepts of measuring change was quantified by correlation coefficients. Please state what you consider as high, moderate, or low correlation/agreement and why. Give a reference for that.

5. P. 10, Present state effect. “The mean correlation between ... were 0.23 and ... 0.41. I had to read this sentence several times until I understood that you calculated the average of the correlations of the 3 scales of Table 3. This does not make sense. I suggest to compare the correlations within each of the scales. Then, the differences may be consistent over all 3 scales. If so, the difference of two correlations could be compared by a statistical test to prove significance within each scale. This would be more interesting than comparison to 0.00 as done in Table 3.
6. Table 3. What is the meaning of the correlation between the transition rating (a change) and the pre- and post-test rating (a status) and how do these results contribute to answer the study’s aim?

7. With respect to the sentence “To our knowledge there is only one study …”(p. 13): Please state the results of that study (Ref. 17) and compare them to your results.

8. To conclude the study, discuss the consequences of the results for clinical work.

Minor essential revisions: none
Discretionary revisions: none

**Level of interest:** An article of importance in its field

**Quality of written English:** Not suitable for publication unless extensively edited

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**
I declare that I have no competing interests.-