Reviewer’s report

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

Version: 1 Date: 29 February 2012

Reviewer: Wilco Emons

Reviewer’s report:

This paper deals with the important topic of measuring person’s change in health outcomes. Measuring change is a challenging topic. This paper compares three approaches (direct, indirect, and quasi-indirect) in an empirical data set. Two retrospective approaches and one based on a pre- posttest design. The paper presents a nice empirical example of how different change methods may produce similar results, but unfortunately I also see some serious limitations in the current study, which I have addressed below.

Major Compulsory Revisions

Research Question & Background & Theory

1. In my view the paper needs more focus and the research questions (including research hypothesis) need further explication. What do you want to learn (or what can we learn from the empirical analysis) about these three measures of change? What are the premises (e.g., is the ideal situation one in which all methods produce exactly the same results)? What are your theoretical expectations/hypotheses?

Methodology:

2. One potential problem with the current research design is that all respondents also completed the scales at baseline. This may yield method bias in the results obtained for both the then-test and direct method because respondents may have matched their answers on the retrospective pre-test to the actual pre-test, sensitized to particular changes, etc. (Actually, one reason for direct or quasi-direct assessment is to circumvent such pre-testing effects). I suggest to comment on this aspect and to motivate to what extent this pre-testing may or may not have limited the findings of the current study. In doing so, it should be incorporated that the results are based on pre- post measurements with a relatively short time interval (3 weeks as mentioned in the discussion).

3. One of the goals of the paper is to analyze the presence or impact of the present state effect on change assessment, but presence and impact are two
different things. Regarding the presence, how do you know whether the effect is present? Thus, what are the signifying characteristics for absence/presence of present-state effects? (This may serve as guidance for the analysis and interpretation of the results). Regarding the impact: one wants to answer the question: “Would we find different estimates of health outcomes if the person’s present state would have been different?” In the current design, the latter question cannot be answered, because present-state is not (and I think cannot be) isolated as a single explanatory variable. I suggest to be more specific about the present state effect (what is it, how can we find it in the current research, etc.).

4. I am not sure whether I fully understood how the direct assessments of change were obtained (see Methods: Analysis Section). It is stated that “The items of the direct method … comprised five categories”. But how many items did you use? What was the content of the items? (Or did you use the items from the three scales, but now using a different response format?) And how did you aggregate the single-item means to an overall rating? More information on the direct change questionnaire is badly needed to fully appreciate the results of the study (e.g., results related to the present-state effects).

Statistics/Results

5. Table 2, results for the SF-36: Why is the sample size (n = 383) for the indirect change score larger than for single pre, retro, or post test scores (i.e., N <= 381)?

Conclusions/Discussion

6. (Table 3) The correlation between indirect and quasi-indirect change measure will be spuriously high because the post-test measure is involved in both change scores. So this high value is artificial and firm conclusions (“high correspondence”) based on it are not to be trusted.

7. Correlation between actual pre-test and retrospective pre-test is indeed high, but the correlation can be positively biased by pre-testing effects (see some of my comments above). I would say that’s another reason not to be overly optimistic.

8. The author(s) state that “the relationship of direct with indirect measurement of change was quite substantial”. I wouldn’t consider correlations in the range .35-.50 (explained variances 10% to 25%) substantial, given that you aim at measuring the same amount of change. So the conclusion needs better substantiation. And, I also noticed that Table 3 does not report the correlation between direct and quasi-indirect change assessments. Why not?

9. “This indicates that the assumed present state effect .. the intended measurement of change”. I cannot follow the argument. How would the comparison of these two correlation values informative about the present-state effect (i.e., if the present-state effect is real, who would the correlation reveal it, and why)? And what do you mean by override?
10. In addition to the previous comment: The discussion on the present state effect is incomplete and hard to understand (at least to me). This was partly because of the wording. For example, what do you mean by saying “these (correlations) are far from claiming to be redundant?” But also because the results themselves were hard to interpret. The correlations between direct ratings and posttest measures were indeed weak, but what does it mean? Why didn’t you look at the correlation between the other two change assessment methods and the post-test measures as a frame of reference? And what are the attenuation effects of reliability (you have a multi-item direct measure, so you have information about the reliability, right)? In short, I think the paper needs an articulated rationale on how to identify the present state effect, which then could be used for a thorough discussion of the effect.

- Minor Essential Revisions

Background:
11. Information about the time-interval between the measurements t0 and t1 should be given in the introduction.

12. If I understood it correctly, the “then-test” involves administering the same scale twice, once asking about the situation t0 and once asking about the situation at t1; right?

13. Two potential biases for the indirect measures are mentioned. I understand the first one, but I don’t see why “the calculation of the delta lies in the hand…” produces bias, nor is an advantage. Isn’t it simply a matter of subtracting two numbers?

14. Sentence “On the negative … of this method”; which method does this refer to?

Section outcomes
15. The outcome measures were selected on psychometrics motivations, but are these also relevant measures from a clinical perspective for the clinical population (i.e., musculoskeletal/cardiac disease)? So you could you say more about the clinical role of these scales (do you expect respondents to change on these variables).

16. I suggest to include the reliabilities (e.g., Lambda 2) of the multiple-item scales used.

Discussion
17. “Quasi-indirect measures … are more efficient … to obtain…”. Why? Isn’t it the same as for indirect measures?

- Discretionary Revisions

No comments.
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 have no competing interests.