Author's response to reviews

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

Authors:

Thorsten Meyer (meyer.thorsten@mh-hannover.de)
Susanne Richter (susanne.richter@drv-nord.de)
Heiner Raspe (heiner.raspe@uksh.de)

Version: 3 Date: 24 January 2013

Author's response to reviews: see over
We want to thank the reviewer and editor for your careful assessment of our manuscript. We have commented on each point raised and hope it provides satisfactory answers to your concerns.

Comments of reviewer

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 page 10 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.).

On page 9 we wrote: “Our re-analysis focused on these three scales because they were the only ones from the original study to apply all three methods of change measurement using the same number of items and featuring equivalent item content. An item of the sleep scale concerning disturbed sleep provides an illustrative example. Patients were asked about the extent to which their sleep was disturbed both before (t0) and after (t1) rehabilitation. At t1, they were also asked either how their problem they possibly had with their sleep being disturbed had changed (direct measurement of change) or to rate the extent to which their sleep had been disturbed at t0 (retrospective pre or then-test).”

For further clarification of this issue we have added another sentence in the analysis section of page 10: “This means that the resulting score in direct measures of change is not a single item rating, as it is often used in transition ratings, but is based on the same number of items as the score calculated in indirect or quasi-indirect measures of change.”

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.

We calculated the ICC for absolute agreement between pre-test × retrospective pre-test scores. Results show that ICCs hardly differ from the rs, therefore the interpretation that has been made with regard to agreement on basis of the Pearson coefficient can be substantiated. The respective values have been added to table 3.
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).

We have also thought about this model of analysis but preferred the Guyatt-approach for mainly two reasons. One reason relates to comparability of results with other studies who have applied the Guyatt-approach. The second reason relates to the violation of principal independence of independent variables, because the indirect measures of change are calculated using the respective post test scores. Therefore, there is an artificial association between these variables which cannot be separated.

For your interest: we have done this analysis as you suggested and did not find a statistical significant additional contribution of the post test score to the transition rating. Still – we believe this analysis to be biased and there decided not to report on it.

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”.

We modified this section and added our analysis of absolute agreement using the ICC (see above). We did not completely remove the idea of this sentence because it serves as an argument why we could not compare the direct measures of change with indirect measures of change by means of ICCs. The respective paragraph now reads:

“The level of agreement between indirect and quasi-indirect as well as direct measures of change (question 1) was calculated by Pearson product-moment correlation coefficients. The status measures on which the indirect and quasi-indirect measures of change were based were on the same scale. The scale of direct measures of change was different from the scales of indirect and quasi-indirect measures of change. Therefore, we calculated the intra-class correlation coefficient (ICC) between pre test and post test measure used for indirect and quasi-indirect measures of change to analyse the level of absolute agreement of both scales, in addition to the Pearson product-moment correlation coefficient. This was not suitable for levels of agreement or direct measures of change with the other measures of change.”

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?

In table 2 we have added the information on the direction of the scale. Therefore the differences in signs of the change scores should be comprehensible.
“Across the board” was a phrase suggested by our native speaker translator. It was meant to say that our results should not be used to uncritically substitute indirect approach to measure change with the quasiindirect approach in different applications of fields (in German we have a similar saying, “durch die Bank”). Since this phrase appears to be equivocal we have modified the sentence slightly.

“Before starting to substitute the quasi-indirect approach of measuring change for the indirect approach in different applications,...”

Additional comments of the editor

... At the end of page 16, you state that you would not expect the status assessment at t0 to bias the assessment at t1, because "most patients have substantial difficulties remembering even important activities and interactions from the admission phase ...". So, here, you are arguing that patients are unable to recall their specific *response(s)* to the assessment at t0. On the other hand, the "then-test" of the quasi-indirect change method is based on the assumption that patients can recall their *state* at t0 quite well. This is indeed a difference, but a subtle one - and one that not everybody would agree with.

This is an interesting point to consider, and our assumption might appear as a contradiction. Therefore, we have added two sentences to clarify our interpretation:

“This does not preclude that the patients are able to remember their health status at the time around their submission. It is reasonable to assume that this information is more general in nature compared to marks on a questionnaire, and it is inextricably linked to a patient’s reason for applying to medical rehabilitation or to his or her perception of the course of illness.”

page 5: Please delete the "and" at the end of the sentence starting with "However, a number of studies have only been able to detect low ...". In the next sentence, please delete the extra comma before the reference number 5.
Done.

page 6: Not sure what caused the "Error!" messages on this page, but please double-check if this is something that you can correct.
We cannot see the error in our file but assume that it is related to the cross-reference number. Therefore, substituted the respective field with a regular number and hope the error disappears.