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: 2 Date: 27 August 2012

Author's response to reviews:

Comments to the reviews

General comment

We want to thank the reviewers of our manuscript for their effort and very instructive comments they provided us. After working through their comments we have decided to make major revisions of our manuscript. Since these changes affected all parts of the manuscript, at some point in the revision process it was not useful to make track changes any more. We have tried to pick up all of the critics, ideas and suggestions and rewritten our manuscript with our original ideas in mind but with a clearer focus. Therefore, we have not explicitly commented on every suggestion made but picked them up in different sections of our revised manuscript. Also, we have asked a professional translator to work through the manuscript after we had made our revisions. This has – in our view –improved the readability of the manuscript markedly and hopefully has reduced ambiguities.

Review #1

One major compulsory revision was concerned with the research question, background and theory. We have focused our manuscript on three research questions that are elaborated on in the background section and explicitly followed in our manuscript. They are concerned with the general question regarding the level of agreement between direct and indirect as well as quasi-indirect and indirect measures of change, and two more specific questions regarding recall bias and the present state bias. The manuscript should have a clearer structure and rationale and we hope we have managed to get our original ideas across in a better way.

We have also picked up the problem of pre-testing effects, i.e. that all respondents also completed the scales at baseline. Since this problem is deeply rooted in our research design, we could only make aware of the problem and discuss how it might have affected our results.
With regard to the present state bias we have followed up the theoretical and empirical work by Guyatt and coworkers. We have changed our approach to analysis for this question and presented new results in the results section.

There was still some ambiguity regarding the direct assessment of change. Therefore, we have tried to make it clear at different parts of the manuscript, how we came up with the scores of the direct assessment of change, including two additional examples. The main point is, that we used items that constituted a scale, and did not use general single items as it is often done with transition ratings. These resulting scales can be analyzed with regard to their reliability. Also, an estimation of effect size can be obtained.

We have also checked our data carefully because of indications for inconsistencies in sample sizes in Table 2. We have found a minor data filter problem and slightly modified Table 2 accordingly.

The conclusion / discussion section has been modified substantively, due to the changes made in the background, analysis and results section. The correlation between direct and quasi-indirect change assessments is not discussed because the study design was set up to randomize patients either to fill in the direct change measurement of the quasi-indirect change measurement, as has been visualized in figure 2.

With regard to the minor essential revisions:

Information about the time-interval between the measurements t0 and t1 has been added. We have also added information to the clinical role of the scales we selected for our analyses. Reliabilities (Cronbach’s alpha) of the scales have been added to the results section. We have also avoided the term “efficient” in our manuscript realizing that there are different uses of the word but left the point we wanted to make in the discussion section, i.e. quasi-indirect changes measures are more economical to obtain compared to indirect changes measures (and therefore, everything else being equal, more efficient for the one who uses or has to pay for obtaining the information).

Review #2

Due to the comments of review #1 we have decided to make our rationale and background more explicit. Therefore, the introduction section has not been shortened. However, we have reduced the information on patient data in table 2 as suggested.

We have put a lot of effort in revising the whole manuscript with regard to clarity of concepts and writing. We hope this effort has been successful. There was a difficult discussion surrounding the best term regarding the title of our manuscript. We tried to avoid the term “correlation” because it is a statistical term and we wanted to emphasize the substantive, not the “technical” message. “Agreement” was not our first choice, because it usually is used to refer to results being identical (like in inter-rater agreement). This situation cannot be achieve in
the comparison of (quasi-)indirect and direct change measurements, because they do not operate on the same scale. Therefore, we used the rather artificial term “correspondence”. However, we realized that in the literature the term “agreement” is also used in a less restrictive sense. Therefore, we have changed our title according to your suggestion and inserted a footnote on our choice of word.

There is one major comment we would disagree with. We believe that it is 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. Direct measures of change were based on the aggregation of a number of transition items relating to the same outcome domain, i.e. they are multi-item scales. They resulted in nicely distributed aggregated responses representing the degree of change in one outcome domain. We agree that the assumption of similar distances of the thresholds has to be proven. However, this holds also true for the pre and post status measurements for which a similar assumption is made. This was not in the realm of our manuscript. We have checked our results by applying the non-parametric Spearman rank correlation coefficient to our bivariate analyses, which did not result in any relevant modifications of our results.

We have picked up the suggestion of using the standardized response mean for the calculation of effect sizes and have changed our results section accordingly. We have also cancelled out the information on mean correlations of the three scales, which made the manuscript more difficult to read and which did not add new information. We do not see at this point well-founded consequences for the clinical work in our manuscript. However, there are consequences for further research, which we comment at the end of the discussion section.