Author's response to reviews

Title: Identification of ambiguities in the 1994 chronic fatigue syndrome research case definition and recommendations for resolution

Authors:

Bart Stouten (bartstouten@wanadoo.nl)

Version: 2 Date: 24 February 2005

Author's response to reviews: see over
Dear editors,

I have revised my manuscript in the light of the reviewer’s comments, with the exception that I have not added confidence intervals for the reported values in Table 1 as was suggested by Referee 1 (see below for details). I made some required changes to the format of my paper. I herewith submit a point-by-point description of the changes made.

Yours faithfully,
Bart Stouten

================================================================================

Referee 1 (Gordon Taylor, 22 November 2004)

Minor Essential Revisions:

I changed “\{S_{\text{min}}, S_{\text{min}}+1, ..., S_{\text{max}}-1, S_{\text{max}}\}” to “\{S_1, S_2, ..., S_{m-1}, S_m\}.” (Since “S_m – S_{m-1}” is not necessarily equal to one, this term now appears explicitly in the denominator of the lower bound, i.e. equation (3) of the revised document. As S_m – S_{m-1} = 1 for all the considered fatigue rating instruments, this does not affect the results in Table 1.)

To interpret the confidence we can ascribe to a certain value it is important to distinguish between the sample that is being studied and the entire population. Properties of the sample, such as the mean age of the individuals in a certain sample, can (generally) be computed with 100% accuracy. This means that the corresponding confidence interval for that particular value, e.g. the sample mean, has zero length. A different situation arises when we use the sample mean as an estimate for the mean of the entire population. Since the sample only constitutes a small part of the population that we study, we can of course not conclude that the population mean is exactly identical to the sample mean. The sample mean serves as an estimate for the true population mean, and the confidence we can ascribe to this estimate can be expressed in terms of a confidence interval.

What I tried to show in my manuscript is that when there are too many items with the maximum score in a particular study, then the results of that particular study must be interpreted with caution. Thus I am interested in the performance of the fatigue rating scales in the particular individual studies, and not so much in the number of items with the maximum score in the entire population. With respect to the samples I studied, the lower bounds that I presented in Table 1 are strict. This means that I can guarantee that extreme scoring occurred at least L-times for that particular sample. (To be precise, I did not take round-off effects in the reported means scores in literature into account in Table 1, but that can be done quite easily and it has only very little effect on the reported lower bounds.)

If we would like to generalize the exact lower bounds that I presented to the entire population, then it would indeed be useful to include confidence intervals. However, two problems arise when we try to do so. Firstly, some of the studies in Table 1 do not report confidence intervals, standard errors, or (sample) standard deviations for the mean fatigue rating scale score Y. For those studies we do not know how well the reported Y reflects the population’s true mean fatigue rating scale score. Since the derived lower bound L depends directly on the reported mean scale score Y, this means that we cannot construct a confidence interval L for these cases, as we do not know the confidence that we can ascribe to Y. Secondly, the lower bound for the percentage of items with the maximum score actually zooms in on upper range of the fatigue rating scales, which means that a small change in the reported fatigue scale score will result in a relatively large change in the presented percentages. Consequently, the confidence intervals for the reported percentages are quite large, and I therefore believe that they are of very limited practical value.

From Table 1 it is clear that there have been several studies in the past where a large number of items reached the extreme end of the scale. I believe that Table 1 in combination with our common sense already provides us enough evidence that there will be many more chronic fatigue syndrome (CFS) studies in the future with abundant extreme scoring if scientists continue to use the “recommended” fatigue rating instruments. I feel that adding relatively large confidence intervals to the presented lower bounds will not alter this view.

To strengthen my argument that the recommended fatigue rating scales will indeed continue to result in abundant extreme scoring in future CFS studies, I would like to mention two recent articles that have been published after I submitted my original manuscript. In a recent randomized controlled trial on the effectiveness of cognitive behavior therapy for CFS that used the fatigue severity subscale of the recommended Checklist...
Individual Strength as a primary outcome, extreme scoring at baseline occurred on at least 56% and 45% of the items in respectively the treatment and the control group (Stulemeijer et al., BMJ 2005;330:14-18). Furthermore, in a study on brain 5-HT1A receptor binding in CFS that used the recommended 11-item continuous Chalder Fatigue scale to measure fatigue, extreme scoring occurred on at least 49% of the items. Noticeably, in the latter study the authors remark that: “However, as in many samples of patients with CFS, there was relatively little variance in the measure of the primary symptom, fatigue, because most patients scored at or near the top of the scale. This ceiling effect makes it more difficult to find significant correlations with any biological measures, added to the subjective nature of symptom reporting” (Cleare et al., Biol Psychiatry 2005;57:239-246).

Referee 2 (John Martin Bland, 3 February 2004)

General:

I rewrote the first section “Derivation” of the Appendix (Page 5 of my original submission) to make it easier to follow.

Minor Essential Revisions:

- Page 2, line 5. Corrected “Thus not matter” to “Thus it does not matter,” and rewrote the last part of the sentence for improved readability.
- Page 2, line –9. Replaced “erroneous conclusion” by “misleading results.”
- Page 2, line –2. Replaced “erroneous conclusion” by “misleading results.”
- Page 5. I added some text to explain the term “item.”
- Page 7, line –2. Corrected “proof” to “prove.”

Formatting

- Provided a separate title page;
- Reformatted the caption of Table 1;
- Added a description of the Author’s contributions;
- Changed the competing interest from “None declared” to “The author declares that he has no competing interests.”