# Challenges to interpreting patient reported outcomes in clinical trials: author rejoinder

## Melanie L. Bell

Department of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, Arizona, USA

*Correspondence to:* Melanie L. Bell, PhD. Department of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona, 1295 N. Martin Ave., P.O. Box 245210, Tucson, Arizona 85724, USA. Email: melaniebell@email.arizona.edu.

It is a pleasure to respond to the editorial by Professor Diane L. Fairclough, who has made foundational contributions to the design and analysis of clinical trials that use patient reported outcomes (PROs) (1). I thank her for adding to the discussion of PRO interpretation, and largely agree with the points raised.

As researchers in the PRO field, one of our greatest challenges is translation and interpretation of PROs into measures that are useful, not only for other PRO researchers, but for clinicians, patients and policy makers. Thus, I agree wholeheartedly with Dr. Fairclough's first point of concern, that many clinicians may not understand what a standardized effect size (SES) is. There is work to be done in education and translation, but I would argue that this work is considerably lessened by transforming our PRO measures to a standardized scale, on a single figure, with values transformed with uniform directionality (favors treatment on one side of the null effect, favors control on the other), as outlined in the original paper (2). A few sentences of interpretation, along with the SESs forest plot should suffice. For example, in addition to showing the forest plot, the PRO researcher could include an explanation such as "SESs are scaled so all effect sizes are on the same scale. Small, medium, and large effect sizes have been suggested as 0.2, 0.5, and 0.8 (3). Important differences for PROs are usually estimated to be an effect size between 0.2 and 0.5" (4). If PRO researchers have investigated the minimum important difference (MID) for one or more of the PROs, these could be marked on the 95% confidence interval (e.g., with an asterisk) for that PRO, or stated in the text (4). Small, medium and large effect sizes have been derived for odds ratios, relative risk and hazard ratios (5,6).

The second issue that Dr. Fairclough raises is the issue of clinical decision making. How does a clinician take information, on a continuous scale, standardized or not, and make decisions to treat or not? But Dr. Fairclough points out that many PROs, such as quality of life or cognitive functioning, are multi-dimensional. The implications here are that one cannot reduce the multiple PROs to a single interpretable number, with an easy-to-make binary decision associated with it. This stems from the complexity of both humans and clinical decision-making, which point to the need for nuanced use of PROs. Others have discussed the problems with dichotomizing information, which include a loss of power and potential for classification bias (7).

PROs are used both as primary outcomes and as secondary or exploratory outcomes. The example that my coauthors and I originally used to demonstrate forest plots and SESs was a randomized controlled trial (RCT) of a web-based brain training program to reduce perceived cognitive impairment following chemotherapy (8). This is an example where PROs made up both primary (perceived cognitive impairment) and secondary outcomes (other dimensions of cognition, quality of life, anxiety/depression, stress, and fatigue). In this case it is likely many readers would be familiar with the concept of SES and the meaning of their magnitudes. Where PROs are used as secondary, tertiary, or exploratory outcomes, for example as supporting
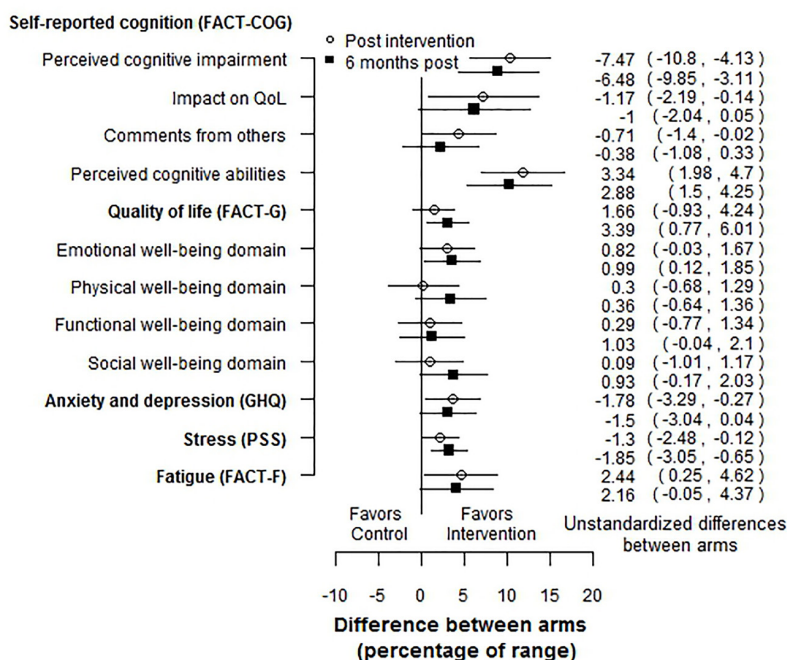
**Figure 1** Results from a RCT of 242 cancer survivors to test a web-based brain training program for reducing self-reported cognitive impairment. The difference between arms as a percentage of each of the PRO scales is shown, along with the unstandardized differences. RCT, randomized controlled trial; PRO, patient reported outcome.

evidence for drug labelling claims (9), readers may be less familiar with PROs, and a standardized scale and explanation may be useful.

Another possible forest plot for multiple PROs would show the differences between arms as a percentage of the PRO scale range. For example, if the scale or subscale had a range of 50, and the control and intervention arms had means of 25 and 30 respectively, the difference would be 5 points, which would be 10 percentage points of the range. An example is shown in *Figure 1*, using the RCT mentioned above (8). This figure clearly shows that RCT gives evidence in favor of the intervention, with small effects, that are slightly attenuated over time. Since the difference in means is divided by a scalar, the p-values are the same as on the absolute scale. This calculation can easily be carried out using a contrast in a mixed model. As with the SES forest plot, we can see that the intervention was effective for most PROs, with effects that attenuated slightly over time. This approach would not be usable for scales based on item response theory (10).

As the importance of the patient voice in research continues to be emphasized, researchers in the field of PROs must continue to work towards improved interpretability of these measures for their readers.

## Footnote

*Conflicts of Interest:* The author has no conflicts of interest to declare.

## References

1. Fairclough DL. Design and Analysis of Quality of Life Studies in Clinical Trials, Second Edition. Boca Raton, FL: Chapman and Hall/CRC, 2010.
2. Bell ML, Fiero MH, Dhillon HM, et al. Statistical controversies in cancer research: using standardized effect size graphs to enhance interpretability of cancer-related clinical trials with patient-reported outcomes. Ann Oncol 2017;28:1730-3.

3.  Cohen J. Statistical Power Analysis for the Behavioral Sciences (2nd Edition). Hillsdale, NJ: Lawrence Earlbaum Associates, 1988.
4.  Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61:102-9.
5.  Olivier J, Bell ML. Effect sizes for 2×2 contingency tables. PLoS One 2013;8:e58777.
6.  Olivier J, May WL, Bell ML. Relative effect sizes for measures of risk. Commun Stat Theory Methods 2017;46:6774-81.
7.  Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ 2006;332:1080.
8.  Bray VJ, Dhillon HM, Bell ML, et al. Evaluation of a Web-Based Cognitive Rehabilitation Program in Cancer Survivors Reporting Cognitive Symptoms After Chemotherapy. J Clin Oncol 2017;35:217-25.
9.  Patrick DL, Burke LB, Powers JH, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. Value Health 2007;10 Suppl 2:S125-37.
10. Hambleton RK, Swaminathan H, Rogers HJ. Fundamentals of item response theory. Newbury Park, CA: Sage Publications, 1991.