Continued Medical Education

Responsiveness analysis: a psychometric consideration in evaluating patient-reported outcome measures

Chinthana Perera[1*] **& Mizaya Cader**[2]

[1]National Dengue Control Unit, Ministry of Health, Sri Lanka; [2]National Programme for Tuberculosis Control and Chest Diseases, Ministry of Health, Sri Lanka

*Correspondence: pcsperera@yahoo.com   https://orcid.org/0000-0001-6067-7345

Introduction

There is a greater need for meaningful interpretation of patient-reported outcomes following interventions, especially health-related quality of life outcomes (HRQoL). Application of relevant and psychometrically sound instruments to measure such outcomes in research and practice assist patients, their family members, clinicians and policy makers to understand the impact of an intervention on respective outcomes (1).

When the psychometric properties, reliability and validity of an instrument are established, it is said to be appropriate to use for research purposes. However, when it is used for evaluation of a change of a variable, it should be sensitive enough or responsive enough to detect the change when patients improve or deteriorate. Responsiveness or sensitivity to change covers an instrument's ability to accurately detect meaningful change when it occurs (2-3). Some researchers indicate that responsiveness is an aspect of validity, not a separate dimension (4).

Measurement of responsiveness

Measurement of the statistical significance of change score either between intervention group and control group, or in the same group as pre- and post-measurements depends on the sample size of the study. With large samples, trivial differences may become statistically significant and with a small sample size, large differences may not become statistically significant. As such, there is a growing concern that assessment of change following an intervention should not solely depend on the statistical significance. Estimating the magnitude of the difference between change scores in both groups, the difference between mean change scores are expressed in standard deviation units with the effect size index (ES) (3). Therefore, effect sizes should report with appropriate statistical test and p values as the number of observations required to detect change has not been estimated with power analysis. Effect size analysis should take as a supplement for statistical testing but not as a substitute (5-6).

There are two broad methods for measuring responsiveness, namely distribution-based methods and anchor-based methods (7-8).

1. Distribution-based methods (8-11)

In this method, the observed change is compared to the statistical properties of the sample, which measures variation attributable to chance. These methods are further classified as group-level and individual-level analyses.

1.1 Group level analysis

- **Cohen D Effect Size (ES)**

Cohen's D ES is calculated by dividing the mean difference of paired measurements of the group by pooled standard deviation (SD) (3).

$$ES = \frac{(X_1 - X_2)}{SD_{pooled}}$$

$$SD_{pooled} = \sqrt{\frac{(SD_{baseline})^2 + (SD_{outcome})^2}{2}}$$

The ES is interpreted by cut-off values recommended by Cohen, which constitute standards for 'trivial' (ES< 0.20), 'small' (ES≥0.20<0.50), medium' (ES≥0.50<0.80), and 'large' (ES≥0.80) (12).

- **Standardized Response Mean (SRM)**

The SRM is calculated by dividing the difference between mean scores (pre and post) by SD of the change score (13). Confidence intervals are calculated by assuming normal distribution (14) using the substitution method (15). Probability of change statistics ((P) is used to interpret the SRM. The P statistic denotes the probability that the instrument detects a change, representing the proportion of subjects whose scores have changed, and ranges from 0.5 (no ability to detect change) to 1 (perfect ability to detect change) (14, 16).

$$SRM = \frac{(X_1 - X_2)}{SD_{change\ scores}}$$

$$\textit{Variance (V) of the SRM is estimated by} = \frac{1}{n} + \frac{SRM^2}{2(n-1)}$$
$$n = sample\ size$$

Approximation of the CI of SRM = SRM $\pm z_{1-\frac{\alpha}{2}} \sqrt{V}$ (14)

Probability of change statistic of SRM is selected from the Z distribution Table (14).

1.2 Individual level analysis

- **Cohen D ES (Individual)**

Cohen's D ES is calculated for all participants individually, by dividing the difference between pre- and post-assessment scores of the instrument by pooled SD. The result is then categorized as ES< -0.5 (large decline), -0.5 to 0.5 (small or medium effect), >0.5 (large improvement). A cut off value of ES> 0.5 is recommended as the threshold for meaningful change. Patients with meaningful improvement more than 0.5 of ES is presented as a proportion (11).

- **Standard Error of Measurement (SEM)**

The SEM is constructed by multiplying the SD of the baseline score by the square root of one minus the reliability of the instrument.

$$SRM = SD \sqrt{1 - (reliability)}$$

Either test-retest reliability or internal consistency reliability (Cronbach's Alpha) is taken to construct the SEM (17-18). Different thresholds of SEM ranging from 1 to 2.77 have been proposed to consider individual level change as statistically meaningful (17-18) while the proportion of individuals with change score more than the set cut-off point is the proportion with statistically meaningful improvement. A responsive measure should have more than 2.5% of the sample to have an increase and more than 2.5% of the sample to have a decrease in their change score greater than 2.77×SEM or 1 SEM, according to the threshold taken (8).

2. Anchor-based methods (7)

In this method, the observed change is related to an external indicator or a criterion of change. It compares the changes in patient-reported outcome scores to other clinically meaningful markers or anchors. They are of two types.

2.1 Longitudinal approach

Correlation of either total scores or change scores of respective instruments (instrument for which responsiveness is assessed) with external standards or criteria will be assessed longitudinally at each time points of the intervention under longitudinal approach (7). Most studies recommend 0.3 to 0.35 as a correlation threshold to define an acceptable association between an anchor and patient-reported outcome change score (1).

2.2 Cross-sectional approach

Cross-sectional approach involves comparison of groups that are different in terms of disease-related criterion to determine clinically meaningful change against anchor criteria. Adhered to this, analysis is carried out to assess whether a significant mean difference of respective instrument under study (instrument for which responsiveness is assessed) exists between patients with and without the disease and between adjacent disease categories (no disease with mild disease, mild disease with moderate disease, etc. e.g. depression) (7). In other words, the change of patient-reported outcome score in the group of people who turned from mild disease to no disease or from moderate disease to mild disease or severe to

moderate disease after the intervention. This change in scores could be considered as the minimum important difference (MID) (19). MID or minimal clinically important difference (MCID) is the smallest change in a treatment outcome that an individual patient would identify as important and which would indicate a change in the patient's management (20-21).

Anchor-based approaches are often easier to interpret for clinical audience than the distribution-based approaches. However, the external anchor chosen must itself be a valid measure of clinical change (22).

References

1. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology* 2008; 61: 102-109.

2. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *Journal of Chronic Diseases* 1987; 40(2): 171-178.

3. Middel B & van Sonderen E. Statistically significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *International Journal of Integrated Care* 2002; 2: e15.

4. Hays RD & Hadorn D. Responsiveness to change: an aspect of validity, not a separate dimension. *Quality of Life Research* 1992; 1: 73-75.

5. Thompson B. If statistical significance tests are broken misused, what practices should supplement or replace them? *Theory & Psychology* 1999; 9(2): 165-181.

6. Thompson B. Editorial policies regarding statistical significance tests: further comments. *Educational Research* 1997; 26(5): 29-32.

7. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *Journal of Clinical Epidemiology* 2003; 56(5): 395-407.

8. Maheswaran H, Weich S, Powell J, Stewart-Brown S. Evaluating the responsiveness of the Warwick Edinburgh Mental Well-Being Scale (WEMWBS): Group and individual level analysis. *Health and Quality of Life Outcomes* 2012; 10: 156.

9. Liang MH. Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. *Medical Care* 2000; 38(9): II84-II90.

10. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *Journal of Chronic Diseases* 1987; 40(2): 171-178.

11. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Medical Care* 2003; 41(5): 582-592.

12. Cohen J. Statistical power analysis for the behavioural sciences. New York: Academic Press; 1977.

13. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopaedic evaluation. *Medical Care* 1990; 28: 632-642.

14. Zou GY. Quantifying responsiveness of quality of life measures without an external criterion. *Quality of Life Research* 2005; 14: 1545-1552.

15. Daly LE. Confidence limits made easy: interval estimation using a substitution method. *American Journal of Epidemiology* 1998; 147: 783-790.

16. Lemieux J, Beaton DE, Hogg-Johnson S, Bordeleau LJ, Goodwin PJ. Three methods for minimally important difference: no relationship was found with the net proportion of patients improving. *Journal of Clinical Epidemiology* 2007; 60(5): 448-455.

17. Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. *Journal of Clinical Epidemiology* 2001; 54: 1204-1217.

18. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *Journal of Clinical Epidemiology* 1999; 52(9): 861-873.

19. Jayadevappa R, Malkowicz SB, Wittink M, Wein AJ, Chhatre S. Comparison of distribution- and anchor-based approaches to infer changes in health-related quality of life of prostate cancer survivors. *Health Services Research* 2012; 47(5): 1902-1925.

20. Scheunemann HJ, Guyatt GH. Commentary – goodbye M(C) ID! Hello MID, where do you come from? *Health Services Research* 2005; 40(2): 593-597.

21. Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). *The Journal of Manual & Manipulative Therapy* 2012; 20(3): 160-166.

22. Eurich DT, Johnson JA, Reid KJ, Spertus JA. Assessing responsiveness of generic and specific health related quality of life measures in heart failure. *Health and Quality of Life Outcomes* 2006; 4: 89.