ELSEVIER

Journal of Clinical Epidemiology 60 (2007) 448–455

Journal of Clinical Epidemiology

Three methods for minimally important difference: no relationship was found with the net proportion of patients improving

Julie Lemieux[a,b,*], Dorcas E. Beaton[c,d], Sheilah Hogg-Johnson[d], Louise J. Bordeleau[b], Pamela J. Goodwin[b]

[a]*Unité de recherche en santé des populations, Centre de recherche, Hôpital du Saint-Sacrement, Quebec, Canada*
[b]*Samuel Lunenfeld Research Institute of the Mount Sinai Hospital, Toronto, Canada*
[c]*Mobility Program Clinical Research Unit and the Martin Family Centre for Arthritis Care and Research at St. Michael's Hospital, Toronto, Canada*
[d]*Institute for Work and Health, Toronto, Canada*

Accepted 21 August 2006

Abstract

Objective: To determine the impact on a responder type analysis of using three published methods to obtain the minimally important difference (MID) on the conclusion of a randomized controlled trial (RCT).

Study Design and Setting: Using data from an RCT of supportive-expressive group therapy (SEGT-intervention) vs. standard care (control) in women with metastatic breast cancer, we measured individual responsiveness to change according to three levels of predefined MID (0.2 SD, 0.5 SD, and 1 standard error of measurement) of the following six validated questionnaires: Profile of Mood States, Impact of Event Scale, Psychosocial Adjustment to Illness Scale, EORTC Quality-of-Life Questionnaire Core-30, Mental Adjustment to Cancer, and a pain visual analog scale. The proportion of women improved by SEGT and the number needed to treat according to three levels of MID were calculated.

Results: There was no consistent difference in the net proportion of women improving with the SEGT vs. control arm according to the three different levels of MID.

Conclusion: The choice between different levels of distribution-based MID did not make an important difference in the net proportion of women improving with the SEGT. Future research should compare MID derived from clinical anchors, in particular patient opinions.

Keywords: Minimally important difference; Responsiveness to change; Psycho-oncology; Health-related quality of life; Supportive-expressive group therapy; Metastatic breast cancer

1. Introduction

Questionnaires used for measurement of health-related quality of life (HRQOL) and psychosocial outcomes in clinical trials must be selected carefully. These questionnaires not only need to be valid and reliable but they also need to be responsive to change in selected outcomes [1]. Responsiveness to change has been defined by de Bruin et al. as ''the ability of a measure to accurately detect change when it has occurred'' [2]. Interpretability refers

Analysis of responsiveness at the group level has been presented in these abstracts. However, analysis of responsiveness at the *individual level*, the subject of the present paper, has never been presented elsewhere.

Lemieux J, Beaton D, Bordeleau L, Hogg-Johnson S, Goodwin P; Mount Sinai Hospital, Toronto, ON, Canada; St Michael's Hospital, Toronto, ON, Canada; Mount Sinai Hospital, Toronto, ON, Canada; Institute for Work and Health, Toronto, ON, Canada. Evaluation of responsiveness to change to psychosocial group therapy (PGT) and to the process of dying of psychosocial and quality-of-life (QOL) measures. American Society of Clinical Oncology Proceedings, 23(16S): 737S, 2005 (abstract 8035) (Poster presentation).

Lemieux J, Beaton D, Bordeleau L, Hogg-Johnson S, Goodwin P; Mount Sinai Hospital, Toronto, ON, Canada; St Michael's Hospital, Toronto, ON, Canada; Mount Sinai Hospital, Toronto, ON, Canada; Institute for Work and Health, Toronto, ON, Canada. Evaluation of responsiveness to change to psychosocial group therapy (PGT) and to the process of dying of psychosocial and quality-of-life (QOL) measures. Canadian Association of Medical Oncologists, Annual Scientific Meeting, Montreal, March 31, 2005 (abstract).

* Corresponding author. Unité de recherche en santé des populations (URESP) Centre de recherche, Hôpital du Saint-Sacrement, 1050 Chemin Ste-Foy, Québec, QC, Canada G1S 4L8. Tel.: 418-682-7511, ext 7518; fax: 418-682-7949.

E-mail address: julie.lemieux@uresp.ulaval.ca (J. Lemieux).

0895-4356/07/$ – see front matter
doi: 10.1016/j.jclinepi.2006.08.006

to the presentation of results so the targeted audience is able to make sense of the change that is presented [3].

Responsiveness is often summarized with the effect size (ES) statistic (mean change divided by the baseline standard deviation [SD]) [4], or standardized response means (mean change divided by the SD of change) [5,6]. These approaches incorporate group level scores and measure group responsiveness. Using the framework of Beaton et al. [7], responsiveness can be analyzed at the group level (change is averaged for the population under study) or at an individual level (change for each individual is analyzed). Individual responsiveness allows interpretation for an individual patient and can be used to define response criterion for a responder analysis in a clinical trial; it enhances the interpretability of the score changes.

Congruent with de Bruin's definition, studies on the ability to detect change are performed in a sample that is known to have changed in some way. Change can be defined in several ways [7]. There is a minimal level of change that can be theoretically captured by an instrument; this level is defined by a one-increment step on one item in the scale. This change is totally dependent on the structure of the scale, and is the lowest threshold that can be interpreted with any meaning; it is not always clinically meaningful. Other definitions of change incorporate measurement error such as the standard error of measurement (SEM, 67% confidence limits) or 95% confidence limits (1.96 SEM) [8–12]. Change less than this amount is difficult to distinguish from variability due to measurement error and is of questionable clinical relevance. Change can also be defined as the lower bound of important change, or the ''minimally important difference'' (MID) [7,13]. The MID has been defined as ''the smallest difference in score in the domain of interest that patients perceive as important, either beneficial or harmful, and which would lead the clinician to consider a change in the patient's management'' [3]. The same amount of change can be interpreted differently according to the perspective taken (patient, clinician, health care system, or the society) or the domain studied (pain vs. appetite) [7].

The use of MID is particularly relevant in psycho-oncology and HRQOL research because the questionnaires used in such research yield data that can be difficult to interpret. For example, what does a change of seven points on a 100-point depression scale mean? Having a known MID would help to guide interpretation, and facilitate evaluation of statistical significance in the context of clinical relevance. MID can be determined using one of two approaches: anchor-based approach or distribution-based approach. Anchor-based methods require that the change in scores be compared to an external measure (e.g., patient's perception of change). Distribution-based methods use the data generated from the questionnaire without reference to an outside measure. Cohen [14] has defined an ES of 0.2, 0.5, and 0.8 as a small, moderate, and large change, respectively. A change score associated with an ES of 0.2 has been considered by Samsa et al. [15] to correspond to the MID in hypertension but this is not always the case in other diseases. Though controversial [16,17], Norman et al. [18] observed that the MID determined using an anchor corresponded to one-half of the baseline SD in their sample. 0.5 baseline SD is equivalent to a change score associated with a 0.5 ES [18]. Finally, 1 SEM, which reflects measurement error, has been shown in some studies to correspond to the MID derived from patients' perceptions of change measured by global rating scales [8,9]. Other work by the same group has suggested that the MID could be closer to 2.3 SEM than 1 SEM [10]. This difference is probably related to the level of clinical significance selected on the global rating perception of change (''good deal better'' vs. ''a little better''), the studied condition (acute vs. chronic disease) and the choice of the reliability coefficient (test–retest vs. Cronbach) [10].

There is no ''gold standard'' for the determination of MID values, but there might be an impact when these values become the criterion for the responder analyses or as defining nodes in a cost effectiveness analysis. Wyrwich et al. [19] suggested using multiple strategies to determine the clinical significance. As a result, we undertook an analysis on the impact on a responder type analysis of selection of three distribution-based methods to determine the MID of six psychosocial and HRQOL questionnaires used in the Breast Expressive-Supportive Therapy study [20]. Our goals were to determine the potential research implications associated with the selection of different levels of MID and to find the questionnaire associated with the highest proportion of patients improved from the supportive-expressive group therapy. Group responsiveness will be addressed separately.

2. Methods

2.1. Data source

Data came from the Breast Expressive-Supportive Therapy study, a Canadian multicenter (seven centers) randomized trial of a weekly supportive-expressive group therapy (SEGT-intervention arm) vs. standard psychosocial care as needed (control arm) in women with metastatic breast cancer [20]. The intervention was standardized and designed to improve mutual, social, and family support, to enhance emotional expressiveness, to integrate change of self and body image, to improve coping skills and doctor–patient relationship, to detoxify death and dying, to develop a life project, and finally to improve HRQOL [21]. A total of 235 women with metastatic breast cancer were recruited between 1993 and 1998 and randomized in a 2:1 ratio in favor of the intervention arm. The primary endpoint was survival and secondary endpoints were psychosocial and HRQOL outcomes. Results of these outcomes have been published; there was no improvement in survival but there