ELSEVIER

Journal of Clinical Epidemiology 63 (2010) 171−179

Journal of Clinical Epidemiology

A new approach to combining clinical relevance and statistical significance for evaluation of quality of life changes in the individual patient

G. Kemmler[a,*], A. Zabernigg[b], K. Gattringer[b], G. Rumpold[a], J. Giesinger[a], B. Sperner-Unterweger[a], B. Holzner[a]

[a]*Department of Psychiatry and Psychotherapy, Innsbruck Medical University, Innsbruck, Tyrol 6020, Austria*
[b]*Department of Internal Medicine, Kufstein County Hospital, Austria*

Accepted 30 March 2009

Abstract

Objective: Empirical investigation of formerly proposed criteria for relevant changes of health-related quality of life (QOL) regarding their use for monitoring changes in the individual patient. Suggestion of a new criterion trying to overcome the drawbacks of former criteria.

Study Design and Setting: QOL data were collected longitudinally in 160 cancer patients receiving chemotherapy at an oncological outpatient unit, giving rise to a total of 975 QOL assessments. QOL was measured using the European Organization on Research and Treatment of Cancer Quality of Life Core Questionnaire. Several formerly suggested criteria of relevant change (distribution based, anchor based) were compared in terms of both prevalence and statistical significance of the resulting relevant changes.

Results: When considering criteria of relevant change suggested in the literature, high proportions (average: 42.3−48.3%) of reputedly relevant changes were found. The majority of these changes (average: 55.8−62.2%) were statistically insignificant. Combination of an increased threshold for clinical relevance with the concept of statistical significance resulted in a more meaningful change criterion.

Conclusion: Formerly recommended thresholds of relevant change in QOL appear to be unduly low when focusing on the individual patient. A modified criterion is therefore suggested for this case. However, more research is needed for validation and refinement of the proposed criterion.

Keywords: Quality of life; Relevant change; Minimal important difference; Minimal detectable change; Reliable change index; EORTC QLQ-C30

1. Introduction

Although in the past two decades quality of life (QOL) research was dominated by studies involving *groups* of patients, there has been a growing interest in the *individual patient* and his or her QOL in recent years. This shift of focus was probably fostered by the new possibilities computer technology offers for data collection, evaluation, and visualization [1−3]. These new, user-friendly computer systems with data entry by the patient allow the routine assessment and monitoring of patient-reported outcomes, including QOL, not only in the highly equipped setting of a clinical trial but also in the daily clinical practice where the focus usually lies on the individual patient [4].

To benefit from the new possibilities, physicians and other medical staff need a certain amount of guidance. This does not only apply for the practical use of the computer system but likewise for the interpretation of the QOL data which, in contrast to clinical variables, are not very familiar to most medical users [5]. One possibility is to provide cut-off values for QOL scores, for example, ''within normal range'' and ''below normal range.'' Although such benchmarks may be helpful for first orientation, the large interindividual variation of QOL scores limits the use of such a classification. However, as QOL data are usually assessed repeatedly over time, an evaluation of an individual's QOL need not be confined to the actual QOL scores but can also make use of the time course. In particular, an evaluation of differences of QOL scores at adjacent time points is of interest. These changes may be classified according to their size and often a partition into three categories will be sufficient for practical purposes: no relevant change, relevant improvement, and relevant deterioration.

Various criteria for defining clinically relevant change have been proposed. These may be broadly classified into two groups: anchor-based and distribution-based criteria

* Corresponding author. Tel.: +43-512-50423689; fax: +43-512-50423628.
E-mail address: georg.kemmler@uki.at (G. Kemmler).

0895-4356/10/$ − see front matter
doi: 10.1016/j.jclinepi.2009.03.016

What is new?

Key message

- When dealing with the quality of life (QOL) of individual subjects rather than groups, the usual concept of minimal important difference (MID) needs reconsideration and possibly redefinition to account for the issue of statistical significance.

What does this paper add?

- When applying the usual MID criteria to the evaluation of QOL changes in individual subjects, over 50% of the reputedly relevant changes proved to be statistically insignificant in a sample of cancer patients completing the European Organization on Research and Treatment of Cancer Quality of Life Core Questionnaire.
- Similar results are to be expected with other samples and instruments, as the above finding is mainly a consequence of the instrument's reliability.
- A substantial increase of the threshold for ''relevant change'' is proposed to resolve this problem.

Implications—what should change now?

- Incorporation of the concept of statistical significance (type one error) appears essential when defining relevant change for individual subjects. Additional inclusion of the type two error would be desirable, but to allow for this more emphasis should be placed on the validation of criteria assessing the clinical significance of QOL changes.

[6−8]. Anchor-based criteria make use of additional information obtained from the patient (who is asked to rate the relevance of perceived changes in health status or well-being) or from the clinical context (e.g., a laboratory or physiological measure or a rating by the clinician). Comparison of this information with the actually measured differences on the QOL scale allows estimation of the so-called *minimal important difference* (MID) [9] or minimal important change, the threshold to be used to distinguish relevant changes from irrelevant ones. Distribution-based methods, on the other hand, use the statistical distribution of the QOL data to derive threshold values for ''relevant'' changes. Probably the most often used distribution-based change measures are multiples of the standard deviation (SD). Measures involving the scale reliability, such as the standard error of measurement (SEM), are also referred to as distribution-based criteria. However, as this type of measure addresses measurement precision rather than clinical relevance its appropriateness as a criterion of relevant change is doubtful [10].

MIDs have been derived for a variety of QOL measures and diagnostic groups, often by combining anchor-based and distribution-based methods. In oncology, MIDs have been established for various instruments of the FACIT system [7,11,12] and for the European Organization on Research and Treatment of Cancer Quality of Life Core Questionnaire (EORTC QLQ-C30) [13,14], among others. Recently efforts have been made to unify the rather confusing multitude of MIDs and two distribution-based parameters have been suggested as rather crude, pragmatic benchmarks: 1/2 SD [15] and 10% of the range of a scale [16].

The change criteria described above and the resulting recommendations for MIDs have been developed primarily for patient groups rather than individual patients. Although some researchers have pointed out that these criteria will basically retain their validity when being used with individual subjects [17,18], others stress the differences between the two types of application [6,19]. In particular, although for patient groups the sample size can, by an appropriate power analysis, be determined in such a way that clinical relevance of a change also entails statistical significance, this is not possible in the case of an individual subject. Rather in this case, the threshold for a significant change is entirely determined by the scale reliability or, more exactly, its SEM. Hence clinical relevance, as defined above, does not imply statistical significance and consequently the change criteria described above should not simply be transferred to the case of $N = 1$ without further consideration.

Up to now, few articles have dealt with criteria suitable for defining clinically relevant QOL changes in the individual subject. An exception is the integrated approach proposed by Crosby et al. [20]. The suggested criterion is defined as the maximum of two thresholds, one governing clinical relevance and the other controlling statistical significance. The first threshold is based on MIDs derived for patient groups, the second is based on significance levels for individual subjects. The authors demonstrated the usefulness of their criterion in a large sample of obesity patients using a QOL scale with rather high reliability (Cronbach's $\alpha = 0.965$). It remains open how the criterion will perform for QOL instruments with lower reliability, as is usually the case for measures used in oncology and many other fields of application.

Wyrwich et al. [21] also considered the evaluation of intraindividual changes in health-related QOL and suggested the use of one SEM as a criterion of relevant change. The authors validated their criterion by means of anchor-based methods and recommended its use both for individual subjects and groups. However, there is very little protection against type one errors as the criterion is based on a two-sided alpha level of about 0.3 (Z-score of 1).

In summary, there seems to be little consensus about the appropriate criteria to be used for evaluating QOL changes in the individual subject. Moreover, there appears to be limited experience how the suggested criteria work in practice. Also, research with oncological patients and cancer-specific QOL measures regarding this issue is scarce.