

# Minimal clinically meaningful differences for the EORTC QLQ-C30 and EORTC QLQ-BN20 scales in brain cancer patients

J. Maringwa<sup>1\*</sup>, C. Quinten<sup>1</sup>, M. King<sup>2</sup>, J. Ringash<sup>3</sup>, D. Osoba<sup>4</sup>, C. Coens<sup>1</sup>, F. Martinelli<sup>1</sup>, B. B. Reeve<sup>5</sup>, C. Gotay<sup>6</sup>, E. Greimel<sup>7</sup>, H. Flechtner<sup>8</sup>, C. S. Cleeland<sup>9</sup>, J. Schmucker-Von Koch<sup>10</sup>, J. Weis<sup>11</sup>, M. J. Van Den Bent<sup>12</sup>, R. Stupp<sup>13</sup>, M. J. Taphoorn<sup>14</sup> & A. Bottomley<sup>1</sup> on behalf of the EORTC PROBE Project and Brain Cancer Group

<sup>1</sup>Quality of Life Department, European Organisation for Research and Treatment of Cancer, Brussels, Belgium; <sup>2</sup>Psycho-oncology Co-operative Research Group, University of Sydney, Sydney, Australia; <sup>3</sup>Department of Radiation Oncology, The Princess Margaret Hospital, University of Toronto, Toronto; <sup>4</sup>Quality of Life Consulting, West Vancouver, Canada; <sup>5</sup>Department of Health Policy and Management, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, USA; <sup>6</sup>Department of Health Care and Epidemiology, School of Population and Public Health, University of British Columbia, Vancouver, Canada; <sup>7</sup>Department of Obstetrics and Gynecology, Medical University Graz, Graz, Austria; <sup>8</sup>Department of Child and Adolescent Psychiatry and Psychotherapy, University of Magdeburg, Magdeburg, Germany; <sup>9</sup>Department of Symptom Research, University of Texas, Houston, USA; <sup>10</sup>Medical Ethics, University of Regensburg, Regensburg; <sup>11</sup>Department of Rehabilitation Psychology, Tumor Biology Center, University of Freiburg, Freiburg, Germany; <sup>12</sup>AZ Rotterdam-Daniel Den Hoed Kliniek, Rotterdam, The Netherlands; <sup>13</sup>Department of Neurosurgery, Multidisciplinary Center for Oncology, University Hospital CHUV, Lausanne, Switzerland; <sup>14</sup>Department of Neurology, Medical Center Haaglanden, The Hague, Netherlands

Received 8 July 2010; revised 20 September 2010; accepted 15 November 2010

**Background:** We aimed to determine the smallest changes in health-related quality of life (HRQoL) scores in the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire core 30 and the Brain Cancer Module (QLQ-BN20), which could be considered as clinically meaningful in brain cancer patients.

**Materials and methods:** World Health Organisation performance status (PS) and mini-mental state examination (MMSE) were used as clinical anchors appropriate to related subscales to determine the minimal clinically important differences (MCIDs) in HRQoL change scores (range 0–100) in the QLQ-C30 and QLQ-BN20. A threshold of 0.2 standard deviation (SD) (small effect) was used to exclude anchor-based MCID estimates considered too small to inform interpretation.

**Results:** Based on PS, our findings support the following integer estimates of the MCID for improvement and deterioration, respectively: physical (6, 9), role (14, 12), and cognitive functioning (8, 8); global health status (7, 4\*), fatigue (12, 9), and motor dysfunction (4\*, 5). Anchoring with MMSE, cognitive functioning MCID estimates for improvement and deterioration were (11, 2\*) and for communication deficit were (9, 7). Estimates with asterisks were <0.2 SD and were excluded from our MCID range of 5–14.

**Conclusion:** These estimates can help clinicians evaluate changes in HRQoL over time, assess the value of a health care intervention and can be useful in determining sample sizes in designing future clinical trials.

**Key words:** anchoring, deterioration, EORTC QLQ-C30, health-related quality of life, improvement, minimal clinically important difference

## Introduction

The increasingly frequent use of patient-reported health-related quality of life (HRQoL) as an outcome in cancer clinical trials over the years implies a greater need for meaningful interpretations of aggregated HRQoL scores. Determining the minimal clinically important difference (MCID) [1] for HRQoL scores from cancer clinical trials is useful to clinicians,

patients, and researchers as a benchmark for assessing the effectiveness of a health care intervention and for determining the sample size in a clinical trial. Benchmarks for interpreting differences between groups cross-sectionally may differ from those for interpreting changes over time within groups [2]. It is important to determine the MCID for HRQoL outcomes because statistical significance based on, e.g. *P* values, does not provide information about the clinical meaningfulness. In large sample sizes, statistically significant results can be obtained when numerical differences in HRQoL change scores are small and not likely to be clinically meaningful. Furthermore, as more and more studies examine the MCIDs for differing

\*Correspondence to: Dr J. Maringwa, Quality of Life Department, European Organisation for Research and Treatment of Cancer, Avenue E. Mounierlaan 83/11, 1200 Brussels, Belgium. Tel: +32-2-774-1633; Fax: +32-2-772-3545; E-mail: john.maringwa@eortc.be

questionnaires and cancers, it will become evident whether it is possible to generalize and adopt one MCID or a set of MCIDs for all questionnaires and patient groups. It will take a large number of such explorations to increase the confidence and familiarity of investigators. Thus, every study contributing to this question is important.

Investigators have relied on two distinct approaches for identifying the MCID: the anchor-based and distribution-based [3]. Anchor-based methods link HRQoL measures to external criteria, either to a known indicator that has clinical relevance [e.g. progression of disease, performance status (PS), etc.] or to patient-derived ratings of change in health [1, 3]. Distribution-based approaches hinge on the statistical features of the HRQoL data. Commonly used approaches include fractions of the standard deviation (SD) [4] of HRQoL scores, the effect size [5], and standard error of measurement (SEM) [6]. Differences of 0.2 SD [4] or 0.3 SD [7] have been used to estimate the MCID. Some authors suggest that 0.5 SD is a reasonable approximation for the MCID [8], although others feel that this is not generalizable [9, 10]. Thresholds of 1 SEM have also been suggested [11]. Other investigators, using data from patients' ratings of their own global change [12] or from patients' comparisons of themselves to others [13], determined that 5%–10% of the instrument range represents a subjectively significant difference or clinically significant change.

The European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire core 30 (EORTC QLQ-C30) assesses HRQoL in cancer patients with 15 scales, each ranging in scores from 0 to 100. The EORTC Brain Cancer Module (EORTC QLQ-BN20), which is intended to supplement the QLQ-C30 when assessing HRQoL, assesses disease symptoms, side-effects of treatment, and some specific psychosocial issues of importance to patients with brain cancer using four scales and seven single items, also ranging in scores from 0 to 100. Anchor-based methods have been used previously to aid the interpretation of QLQ-C30 scores [12, 14]. Using global ratings of change as the anchor, Osoba et al. [12] suggested that in patients with breast and small-cell lung cancer, changes in scores of 5–10 represented a small difference; 10–20 represented a moderate difference while those above 20 represented large differences. Using a variety of clinical classifications as anchors, King [14] came to similar findings after collating results from various studies and various cancer sites. Based on these two studies, mean differences of 10 points or more are widely viewed as being clinically significant when interpreting the results of randomized clinical trials that use the QLQ-C30 [15]. However, the evidence is not clear that a 10-point threshold is applicable to each of the 15 QLQ-C30 scales [15]. Furthermore, it has not yet been established whether the same thresholds apply to improvement and deterioration in HRQoL scores. These thresholds may also vary across patient groups. Additional empirical investigation of the size and patterns of MCIDs across domains of the QLQ-C30 in specific patient groups is therefore justified. No MCIDs have been determined yet for either the QLQ-C30 in brain cancer or the QLQ-BN20.

The main focus of this study was to determine the change in selected QLQ-C30 and QLQ-BN20 scales, which corresponds to the MCID for improvement and deterioration in HRQoL for brain cancer patients. Identification of MCIDs was carried out

using two clinical anchors: change in physician-rated World Health Organisation (WHO) PS and changes in the minimal state examination (MMSE).

## patients and methods

### the EORTC QLQ-C30 and the QLQ-BN20

The QLQ-C30 contains both single- and multi-item scales. Of the 30 items, 24 aggregate into nine multi-item scales representing various HRQoL dimensions: five functioning scales (physical, role, emotional, cognitive and social), three symptom scales (fatigue, pain and nausea), and one global measure of health status. The remaining six single-item scales assess symptoms: dyspnea, appetite loss, sleep disturbance, constipation and diarrhea, and the perceived financial impact of the disease treatment. High scores indicate better HRQoL for the global health status and functioning scales, and worse HRQoL for the symptom scales. The QLQ-BN20 contains 20 items, 13 of which aggregate into four scales assessing future uncertainty, visual disorder, motor dysfunction (MD), and communication deficit. The remaining single items assess other disease symptoms (e.g. headaches and seizures) and treatment toxic effects (e.g. hair loss) [16]. For all these scales, a higher score represents worse HRQoL.

### description of the data

Two closed EORTC randomized controlled clinical trials (RCTs) enrolling in total 941 high-grade glioma patients were jointly analyzed. Trial 1 reported by Stupp et al. [17] compared radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone, enrolled 573 patients and used the EORTC QLQ-C30 version 3. Trial 2 reported by van den Bent et al. [18] was an RCT comparing radiation therapy with and without combination chemotherapy, which enrolled 368 patients and used version 2 of the EORTC QLQ-C30. These two versions of the QLQ-C30 differ only in the response options for the items in the physical and role functioning domains. Version 2 uses a binary (no, yes) scale and version 3 uses a 4-point scale ranging from 'not at all' to 'very much' [19]. Analysis pertaining to physical and role functioning scales was restricted to Trial 1, which used version 3, the current version [19]. In both trials, HRQoL was measured as a secondary end point at baseline, during treatment, and on several follow-up occasions after the end of treatment.

### the anchors and selection of QLQ-C30 scales

The anchor-based approach to developing MCIDs requires an independent standard or 'anchor' that is itself interpretable and at least moderately correlated with the instrument being explored [20]. The WHO PS and the MMSE were chosen as the clinical anchors against which changes in selected scales of the QLQ-C30 and QLQ-BN20 would be calibrated. These anchors are clearly definable, understandable, and are commonly used by clinicians in assessment of cancer patients and could therefore help guide interpretation of HRQoL scores. The values for WHO PS range from 0 (no symptoms of cancer) to 4 (bedbound). Changes in PS were categorized into three groups: deterioration (PS worsened by one category), no change (PS stayed the same), and improvement (PS improved by one category). The MMSE [21] is a test with a 30-point maximum score, which is used to screen for cognitive impairment. Any score  $\geq 25$  points (out of 30) is effectively normal. Changes in MMSE were grouped as: deterioration (MMSE worsened by 4 or 5 points), no change (MMSE changed by 3 or less points), and improvement (MMSE improved by 4 or 5 points). Changes in MMSE of 4 or more points have been considered in the literature as clinically significant [22, 23]. In this study, changes in MMSE of 6 or more points were viewed as rather too large for the purpose of determining the MCID and were therefore excluded from the analysis, as were the changes in PS of two or more categories in order not to overestimate the MCID.

We conducted MCID analyses only for those scales from the QLQ-C30 and QLQ-BN20 that were expected to be clinically related to the anchors and had a correlation of at least 0.30 with the anchors at baseline [24]. Scales from the QLQ-C30 that were suitable for anchoring against PS, together with the respective correlation coefficients with the anchor were: cognitive functioning (CF,  $-0.31$ ), physical functioning (PF,  $-0.40$ ), role functioning (RF,  $-0.40$ ), global health status (GHS,  $-0.32$ ), and fatigue (FA,  $0.34$ ). The MD scale of the QLQ-BN20 had a correlation of  $0.41$  with PS and was also anchored against PS. Only two scales were suitable for anchoring with the MMSE: cognitive functioning (CF,  $0.35$ ) from the QLQ-C30 and communication deficit (CD,  $0.41$ ) from the QLQ-BN20.

### analysis

Our focus was on changes in individual HRQoL scores of patients over time, anchoring by changes in WHO PS and changes in the MMSE, respectively. Separate analyses were conducted for these two anchors. For each analysis, patients with data on an anchor and HRQoL scores at 2 or more time points were included. The two points furthest apart in time, denoted by  $T_1$  and  $T_2$ , provided a better chance of observing changes in HRQoL scores and were therefore used for analysis.

Changes in the anchor values and HRQoL scores between  $T_1$  and  $T_2$  were calculated for each patient. Each individual patient's HRQoL change score was assigned to one of the three 'clinically meaningful' categories, as defined a priori by the anchors, e.g. 'improvement', 'no change', and 'deterioration' groups for PS, similarly for MMSE. Note that patients could be categorized differently between the anchor and HRQoL measures, e.g. some patients could be classified as having not changed according to the anchor yet their change in HRQoL was not 0, and if this occurred systematically in one direction, the mean HRQoL change for this category would deviate from the expected value of 0. Therefore, to control for the amount of change in HRQoL that occurred to patients who did not change according to the anchor, we obtained estimates of the MCID by calculating the difference in mean HRQoL change between adjacent categories [11], i.e. 'improvement' versus 'no change' and 'no change' versus 'deterioration'. Ninety-five percent confidence intervals (CI) for the differences in mean of change scores between adjacent categories (the MCID) were calculated.

The association between HRQoL scores and anchor values and between changes in the anchor values and changes in the HRQoL scale were quantified by the Spearman's rank correlation coefficient.

For comparison purposes, four distribution-based approaches were applied: 0.2 SD, 0.3 SD, 0.5 SD, and the SEM. Each of these statistics was calculated twice, once with an estimate of SD from  $T_1$  and again with an estimate of SD from  $T_2$ . The calculation of the SEM also requires an estimate of test-retest reliability, which we obtained from previous studies for the QLQ-BN20 [16] and the QLQ-C30 [25]. Our MCID results were also compared with the 5%–10% range of the instrument [13].

### results

Table 1 summarizes selected baseline demographic and clinical characteristics in the combined sample of all the 941 patients from the two trials.

Descriptive statistics summarizing the distributions of HRQoL scores at baseline are given in Table 2. The distributions for PF, RF, CF, MD, CD, and FA were skewed, with a predominance of good functioning and communicating ability, and low fatigue and MD, while GHS appeared reasonably symmetrical.

Fifteen patients improved their MMSE by 6 or more points and 33 patients deteriorated by 6 or more points and were

**Table 1.** Selected baseline demographic and clinical characteristics of the patients

| Descriptive statistics      | Number (%) |
|-----------------------------|------------|
| Gender                      |            |
| Male                        | 572 (61.0) |
| Female                      | 368 (39.0) |
| Performance status          |            |
| 0                           | 357 (38.0) |
| 1                           | 448 (48.0) |
| 2                           | 131 (14.0) |
| Country                     |            |
| Australia                   | 3 (0.3)    |
| Austria                     | 32 (3.4)   |
| Belgium                     | 57 (6.1)   |
| Canada                      | 168 (17.8) |
| Finland                     | 6 (0.6)    |
| France                      | 153 (16.3) |
| Germany                     | 79 (8.4)   |
| Hungary                     | 10 (1.1)   |
| Israel                      | 14 (1.5)   |
| Italy                       | 81 (8.6)   |
| Poland                      | 3 (0.3)    |
| Slovenia                    | 3 (0.3)    |
| Spain                       | 19 (2.0)   |
| Sweden                      | 10 (1.1)   |
| Switzerland                 | 52 (5.5)   |
| The Netherlands             | 208 (22.1) |
| UK                          | 43 (4.6)   |
| Age ( $n = 941$ )           |            |
| Mean (years)                | 52         |
| Interquartile range (years) | 45–61      |

excluded from further MMSE analyses. Excluded from the PS analyses were two patients whose PS deteriorated by two categories and 51 patients whose PS improved by two or more categories.

The cross-sectional correlations of HRQoL measures with PS were generally moderate, ranging in absolute value from 0.31 to 0.51 (Table 2). The correlations between changes in HRQoL scores and changes in PS were generally weak, ranging from 0.13 to 0.24 in absolute value. Cross-sectional correlations between MMSE and CF at  $T_1$  and  $T_2$  were 0.30 and 0.34, respectively, while for changes in CF and MMSE, the correlation was 0.24. Cross-sectional correlation between MMSE and CD at both times was  $\sim 0.30$  and for the changes, the correlation was 0.15.

The mean change scores for the selected QLQ-C30 and QLQ-BN20 scales and corresponding differences between adjacent categories are presented in Tables 3 and 4, anchored by PS and MMSE, respectively.

MCID estimates anchored by PS are given in Table 3. For illustration, the first difference in PF mean change of adjacent categories is obtained as  $4.3 - (-1.3) = 5.6$  PF units, and the second is calculated as  $-1.3 - (-9.8) = 8.5$  PF units, providing the MCID estimates for improvement and deterioration, respectively. The 95% CI for PS-anchored MCIDs generally did not include zero, suggesting statistically significant differences

**Table 2.** Summary statistics for HRQoL scales at baseline, cross-sectional correlation estimates for HRQoL scores with PS, and correlations between HRQoL change scores with changes in the anchor

|     | HRQoL summary statistics at baseline |        |                |                | Cross-sectional correlation (HRQoL and PS) |                | Correlation between changes in HRQoL and changes in PS |
|-----|--------------------------------------|--------|----------------|----------------|--|----------------|--|
|     | Mean (SD)                            | Median | Q <sub>1</sub> | Q <sub>3</sub> | T <sub>1</sub>                             | T <sub>2</sub> |  |
| PF  | 78.8 (25.0)                          | 86.7   | 66.7           | 100            | -0.40                                      | -0.51          | -0.21  |
| RF  | 60.7 (33.8)                          | 66.7   | 33.0           | 100            | -0.41                                      | -0.47          | -0.21  |
| CF  | 71.5 (27.6)                          | 83.3   | 50.0           | 100            | -0.35                                      | -0.39          | -0.19  |
| GHS | 62.1 (22.8)                          | 66.7   | 50.0           | 83.3           | -0.33                                      | -0.39          | -0.13  |
| FA  | 36.3 (24.8)                          | 33.3   | 22.2           | 55.6           | 0.31                                       | 0.43           | 0.24   |
| MD  | 17.6 (22.5)                          | 11.1   | 0.0            | 22.2           | 0.42                                       | 0.46           | 0.14   |
| CD  | 17.6 (24.7)                          | 11.1   | 0.0            | 33.3           |  |                |  |

Q<sub>1</sub> and Q<sub>3</sub> are the lower and upper quartiles. CD was not anchored against PS and correlations with PS are therefore not presented.

HRQoL, health-related quality of life; PS, performance status; PF, physical functioning; RF, role functioning; CF, cognitive functioning; GHS, global health status; FA, fatigue; MD, motor dysfunction; CD, communication deficit; SD, standard deviation.

between the mean change of adjacent categories except for the 'improvement' and 'no change' comparison for PF and MD and the 'no change' and 'deterioration' comparison for GHS. In the MCID estimates for CF and CD anchored by MMSE (Table 4), statistically significant differences were noted only between the 'improvement' and 'no change' groups.

Table 5 presents distribution-based MCID estimates for comparison with anchor-based estimates in Tables 3 and 4. Since the results for T<sub>1</sub> and T<sub>2</sub> were very similar, only the results at T<sub>1</sub> are reported.

## discussion

The aim of our analysis was to determine the magnitude of difference in scores on selected QLQ-C30 and QLQ-BN20 scales that represents the MCID in brain cancer patients. We used WHO PS as a clinical anchor to determine the MCID estimates for QLQ-C30 scales: physical, role, and cognitive functioning, global health status, fatigue and for MD from the QLQ-BN20. The MMSE was used as an anchor for the QLQ-C30 cognitive functioning scale and the QLQ-BN20 communication deficit scale. To further clarify the issue of MCIDs, the findings from the anchor-based approach were compared with selected distribution-based techniques.

Our analysis combined data from two versions of the QLQ-C30 where one version used a 4-point scale and the other a binary scale. The two versions differ only in the response options for items in the physical and role functioning domains. The 4-point and binary scales are not comparable, and therefore, combining data for the physical and role functioning domains may not be appropriate. In light of this, we restricted analysis of physical and role functioning domains only to the most recent version of the questionnaire; the one that uses the 4-point scale.

In general, the mean changes in HRQoL within each anchor-defined category were in the expected direction. Patients classified as having improved by the anchor tended to have better HRQoL scores, and those classified as deteriorated by the anchor tended to have worse HRQoL scores, while those classified as not changed by the anchor generally did not appear to have changed in their HRQoL (i.e. mean change in HRQoL was roughly 0).

Our results suggest that the MCID estimates for improvement and for deterioration tend to vary across the selected QLQ-C30 and QLQ-BN20 scales. However, we found no clear indication that the MCID for improvement was systematically larger or smaller than the MCID for deterioration. This is in contrast to a number of studies [12, 13, 26] that have found that MCID estimates for deterioration were larger than those for improvement. Further investigation, if possible with other anchors, is therefore recommended.

In general, the anchor-based MCID estimates tended to be larger than the 0.2 SD estimates, smaller than the 0.5 SD estimates, and closer to both the 0.3 SD and SEM. This provides further evidence that the 0.5 SD may represent a 'medium' effect size [5], whereas the 1 SEM [6] and 0.3 SD [7] may approximate a threshold for defining the MCID. The range of anchor-based MCID estimates of 5–14 points is in line with the 5%–10% range of the instrument (5–10 points) [12, 13], which was considered as a subjectively significant difference or clinically significant change.

We acknowledge as a limitation that the observed correlations between the anchors and HRQoL scores were not strong. It may therefore be argued that such anchors may not be used for the particular scales. Other studies [12, 13] have also found only moderately strong correlations of the anchors with the HRQoL scores; the reason(s) are unknown. For interpretation, it could be recommended to augment the anchor-based MCID estimates with results from one of the distribution-based approaches by considering only those anchor-based MCID estimates at least equal to 0.2 SD [11], which is a 'small effect' [5]. This would then imply that the MCID estimates for deterioration for global health status (Table 3) and cognitive functioning (based on MMSE, Table 4) and that for improvement for MD (Table 3) will not be recommended as plausible MCID estimates.

In an anchor-based approach, it is critical that the anchors be understandable and clinically significant. Changes in PS are accepted as clinically significant by most oncologists, and PS criteria is frequently used to determine eligibility for a given treatment or clinical trial. Furthermore, the changes that we are calling MCIDs are based on the definitions and clinical anchors that we have applied, e.g. a one category change on a 5-point WHO PS scale. Had we used a one-category change in the

**Table 3.** Performance status—mean (SD) of HRQoL change scores in the three anchor-defined groups and the MCID (95% CI) between adjacent categories

| Scale | Improved by one category, n (32–85) | No change, n (168–324) | Deteriorated by one category, n (67–163) | MCID (95% CI)   |                       |
|-------|-------------------------------------|------------------------|--|---|-----------------------|
|       |                                     |                        |  | Difference in mean change between adjacent categories |                       |
|       |                                     |                        |  | Improvement   | Deterioration         |
| PF    | 4.3 (25.6)                          | −1.3 (20.3)            | −9.8 (23.7)                              | 5.6 (−2.4 to 13.7)                                    | 8.5 (2.4 to 14.5)*    |
| RF    | 18.7 (28.8)                         | 4.4 (35.1)             | −7.9 (32.3)                              | 14.3 (1.5 to 27.1)*                                   | 12.3 (2.6 to 21.8)*   |
| CF    | 6.8 (26.3)                          | −0.8 (25.1)            | −9.0 (27.7)                              | 7.6 (1.3 to 14.7)*                                    | 8.2 (3.2 to 14.1)*    |
| GHS   | 8.8 (28.4)                          | 1.5 (24.1)             | −2.0 (22.3)                              | 7.3 (1.0 to 13.6)*                                    | 3.5 (−1.0 to 8.0)     |
| FA    | −12.8 (26.3)                        | −0.4 (26.6)            | 8.5 (27.7)                               | −12.4 (−19.0 to −5.9)*                                | −8.9 (−14.0 to −3.7)* |
| MD    | −4.5 (24.7)                         | −0.2 (20.0)            | 5.0 (20.3)                               | −4.3 (−9.6 to 1.1)                                    | −5.2 (−9.0 to −1.3)*  |

PF and RF are based on trial 1 only. The number of patients in the anchor-defined groups varies by the HRQoL scale and is therefore presented as a range of values for all the scales. Differences that are statistically significant are indicated by asterisk. Difference in mean change refers to the difference in mean of HRQoL change scores between the 'Improvement' and 'no change' (Improvement) and between the 'no change and 'deterioration' (Deterioration). HRQoL, health-related quality of life; MCID, minimum clinically important difference; CI, confidence interval; PF physical functioning; RF, role functioning; CF, cognitive functioning; GHS, global health status; FA, fatigue; MD, motor dysfunction.

**Table 4.** MMSE—mean (SD) of HRQoL change scores in the three anchor-defined groups and the MCID (95% CI) between adjacent categories

| Scale | Improved by 4 or 5 points, n (26) | No change, n (371) | Deteriorated by 4 or 5 points, n (23) | MCID (95% CI)   |                     |
|-------|-----------------------------------|--------------------|---------------------------------------|---|---------------------|
|       |                                   |                    |                                       | Difference in mean change between adjacent categories |                     |
|       |                                   |                    |                                       | Improvement   | Deterioration       |
| CF    | 9.0 (22.7)                        | −1.8 (25.1)        | −3.6 (28.0)                           | 10.8 (0.9 to 20.8)*                                   | 1.8 (−8.9 to 12.4)  |
| CD    | −5.7 (19.2)                       | 3.4 (20.2)         | 10.4 (21.7)                           | −9.1 (−16.7 to −1.5)*                                 | −7.0 (−15.6 to 1.5) |

Differences that are statistically significant are indicated by asterisk. Difference in mean change refers to the difference in mean of HRQoL change scores between the 'Improvement' and 'no change' (Improvement) and between the 'no change and 'deterioration' (Deterioration).

MMSE, mini-mental state examination; HRQoL, health-related quality of life; MCID, minimum clinically important difference; CI, confidence interval; CF, cognitive functioning; CD, communication deficit.

**Table 5.** Distribution-based MCID estimates at T<sub>1</sub>

| Scale | Distribution-based method |        |        |      |
|-------|---------------------------|--------|--------|------|
|       | 0.2 SD                    | 0.3 SD | 0.5 SD | SEM  |
| PF    | 5.1                       | 8.6    | 12.9   | 7.7  |
| RF    | 6.7                       | 11.1   | 16.7   | 14.2 |
| CF    | 5.3                       | 8.9    | 13.3   | 11.3 |
| GHS   | 4.4                       | 7.4    | 11.0   | 9.4  |
| FA    | 4.8                       | 8.0    | 12.1   | 10.0 |
| MD    | 4.5                       | 7.4    | 11.1   | 10.0 |
| CD    | 4.5                       | 7.5    | 11.2   | 13.0 |

SD, standard deviation; SEM, standard error of measurement; PF physical functioning; RF, role functioning; CF, cognitive functioning; GHS, global health status; FA, fatigue; MD, motor dysfunction; CD, communication deficit.

10-point Karnofsky performance scale, our results would probably differ. In other words, the MCIDs we have found are not 'absolute' but are 'relative' to the clinical anchors we used.

Another limitation with our study is the limited availability of other appropriate anchors. Being retrospective in nature, our analysis was restricted to use only WHO PS and MMSE, which were deemed credible anchors for the selected scales. Determination of the MCID based on a number of different anchors would be the preferred approach.

In conclusion, our findings provide estimates of MCIDs for brain cancer patients when using the EORTC QLQ-C30. The estimates generally agree with the estimates of 5–10 units of the QLQ-C30 scales we considered and as proposed by Osoba et al. [12] and King [14]. These estimates can be useful to clinicians to determine the proportion of patients benefiting from some treatment. The estimates could also be used as guidance for classification of patients by changes in HRQoL and symptoms over time. Furthermore, the estimates may be useful in sample size determination and design for future clinical trials.

## acknowledgements

We thank the EORTC Clinical Brain Group and their clinical investigators, and all the patients who participated in these trials.

## funding

This study was funded by an unrestricted academic grant from the Pfizer Foundation.

## disclosure

The authors declare no conflicts of interest.

## references

1. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; 10: 407–415.
2. King MT, Stockler MS, Cella D et al. Meta-analysis provides evidence-based effect sizes for a cancer-specific quality of life questionnaire, the FACT-G. *J Clin Epidemiol* 2010; 63: 270–281.
3. Lydick F, Epstein RS. Interpretation of quality of life changes. *Qual Life Res* 1993; 2: 221–226.
4. Samsa G, Edelman D, Rothman ML et al. Determining clinically important differences in health status measures: a general approach with illustration to the Health Utilities Index Mark II. *Pharmacoeconomics* 1999; 15: 141–155.
5. Cohen J. *Statistical Power Analysis for the Behavioural Sciences*. New York: Academic Press 1988.
6. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003; 56: 395–407.
7. Eton DT, Cella D, Yost KJ et al. A combination of distribution- and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. *J Clin Epidemiol* 2004; 57: 898–910.
8. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003; 41: 582–592.
9. Beaton DE. Simple as possible? Or too simple? Possible limits to the universality of the one-half standard deviation (comment). *Med Care* 2003; 41: 593–596.
10. Farivar SS, Kiu H, Hays RD. Another look at the half standard deviation estimate of the minimally important difference in health-related quality of life scores. *Expert Rev Pharmacoecon Outcomes Res* 2004; 4(5): 521–529.
11. Cella D, Eton DT, Lai J et al. Combining anchor and distribution-based methods to derive minimal clinically important differences on the Functional Assessment of cancer therapy (FACT) anemia and fatigue scales. *J Pain Symptom Manage* 2002; 24: 547–561.
12. Osoba D, Rodrigues G, Myles J et al. Interpreting the significance of changes in health related quality-of-life scores. *J Clin Oncol* 1998; 16: 139–144.
13. Ringash J, O'Sullivan B, Bezjak A, Redelmeier DA. Interpreting clinically significant changes in patient-reported outcomes. *Cancer* 2007; 110(1): 196–202.
14. King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. *Qual Life Res* 1996; 5: 555–567.
15. Cocks K, King MT, Velikova G et al. Quality, interpretation and presentation of European Organization for Research and Treatment of Cancer quality of life questionnaire core 30 data in randomized controlled trials. *Eur J Cancer* 2008; 44: 1793–1798.
16. Osoba D, Aaronson NK, Muller M et al. The development and psychometric validation of a brain cancer quality of life questionnaire for use in combination with general cancer-specific questionnaires. *Qual Life Res* 1996; 5: 139–150.
17. Stupp R, Hegi ME, Mason WP et al. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol* 2009; 10: 459–466.
18. van den Bent MJ, Carpentier AF, Brandes AA et al. Adjuvant procarbazine, lomustine, and vincristine improves progression-free survival but not overall survival in newly diagnosed anaplastic oligodendrogliomas and oligoastrocytomas: a randomized European Organisation for Research and Treatment of Cancer phase III trial. *J Clin Oncol* 2006; 24: 2715–2722.
19. Fayers P, Aaronson N, Bjordal K et al. *EORTC QLQ-C30 Scoring Manual*. 3rd edition. Brussels, Belgium: EORTC Quality of life Study Group 2001.
20. Guyatt GH, Osoba D, Wu AW et al. Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 2002; 77: 371–383.
21. Folstein MF, Folstein SE, McHugh PR. Mini-mental state: a practical method for grading the state of patients for the clinician. *J Psychiatr Res* 1975; 12: 189–198.
22. Brown PD, Buckner JC, O'Fallon JR et al. Effects of radiotherapy on cognitive function in patients with low-grade glioma measured by the Folstein Mini-Mental State Examination. *J Clin Oncol* 2003; 21: 2519–2524.
23. Tangalos EG, Smith GE, Ivnik RJ et al. The Mini-Mental State Examination in general medical practice: clinical utility and acceptance. *Mayo Clin Proc* 1996; 71: 829–837.
24. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008; 61: 102–109.
25. Hjermstad MJ, Fossa SD, Bjordal K, Kaasa S. Test/retest study of the European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire. *J Clin Oncol* 1995; 13: 1249–1254.
26. Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores: differences between improvement and worsening. *Qual Life Res* 2002; 11: 207–221.