Establishing anchor-based minimally important differences for the EORTC QLQ-C30 in glioma patients

Linda Dirven, Jammbe Z. Musoro, Corneel Coens, Jaap C. Reijneveld, Martin J. B. Taphoorn, Florien W. Boele, Mogens Groenvold, Martin J. van den Bent, Roger Stupp, Galina Velikova, Kim Cocks, Mirjam A. G. Sprangers, Madeleine T. King, Hans-Henning Flechtner, and Andrew Bottomley on behalf of the EORTC Brain Tumor Group and Quality of Life Groups

Department of Neurology, Leiden University Medical Center, Leiden, the Netherlands (L.D., M.J.B.T.); Department of Neurology, Haaglanden Medical Center, The Hague, the Netherlands (L.D., M.J.B.T.); European Organisation for Research and Treatment of Cancer (EORTC), Brussels, Belgium (J.Z.M., C.C., A.B.); Department of Neurology & Brain Tumor Center, Amsterdam University Medical Centers, location VU University Medical Center, Amsterdam, the Netherlands (J.C.R.); Leeds Institute of Medical Research at St James's, St James's University Hospital, Leeds, UK (F.W.B., G.V.); Faculty of Medicine and Health, Leeds Institute of Health Sciences, University of Leeds, Leeds, UK (F.W.B.); Departments of Public Health and Palliative Medicine, University of Copenhagen and Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark (M.G.); Brain Tumor Center, Erasmus MC Cancer Institute, Rotterdam, the Netherlands (M.J.V.B.); Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA (R.S.); Adelphi Values, Bollington, Cheshire, UK (K.C.); Department of Medical Psychology, Cancer Center Amsterdam, Amsterdam University Medical Centers, location Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands (M.A.G.S.); Faculty of Science, School of Psychology, University of Sydney, Sydney, NSW, Australia (M.T.K.); Clinic for Child and Adolescent Psychiatry and Psychotherapy, University of Magdeburg, Magdeburg, Germany (H.F.)

Corresponding Author: Linda Dirven, PhD, Department of Neurology, Leiden University Medical Center, P.O. Box 9600, 2300 RC Leiden, the Netherlands (l.dirven@lumc.nl).

Abstract

Background. Minimally important differences (MIDs) allow interpretation of the clinical relevance of health-related quality of life (HRQOL) results. This study aimed to estimate MIDs for all European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire Core 30 (QLQ-C30) scales for interpreting group-level results in brain tumor patients.

Methods. Clinical and HRQOL data from three glioma trials were used. Clinical anchors were selected for each EORTC QLQ-C30 scale, based on correlation (>0.30) and clinical plausibility of association. Changes in both HRQOL and the anchors were calculated, and for each scale and time period, patients were categorized into one of the three clinical change groups: deteriorated by one anchor category, no change, or improved by one anchor category. Mean change method and linear regression were applied to estimate MIDs for interpreting within-group change and between-group differences in change over time, respectively. Distribution-based methods were applied to generate supportive evidence.

Results. A total of 1687 patients were enrolled in the three trials. The retained anchors were performance status and eight Common Terminology Criteria for Adverse Events (CTCAE) scales. MIDs for interpreting within-group change ranged from 4 to 12 points for improvement and –4 to –14 points for deterioration. MIDs for between-group difference in change ranged from 4 to 9 for improvement and –4 to –16 for deterioration. Most anchor-based MIDs were closest to the 0.3 SD distribution-based estimates (range: 3-10).

Conclusions. MIDs for the EORTC QLQ-C30 scales generally ranged between 4 and 11 points for both within-group mean change and between-group mean difference in change. These results can be used to interpret QLQ-C30 results from glioma trials.

Key Points

1. Minimally important differences for EORTC QLQ-C30 scales ranged between 4 and 11 points in glioma.

2. Estimate for each HRQOL scale can be used to better interpret group-level results in glioma trials.

Importance of the Study

Minimally important differences (MIDs) allow interpretation of the clinical relevance of health-related quality of life (HRQOL) results. Typically, a score of >10 points was considered clinically relevant for all scales. More recent work, however, suggests that these MIDs are too simplistic as they do not distinguish between domain scales, direction of change (ie, improvement or deterioration), and disease sites. In this study, we estimated MIDs for all EORTC QLQ-C30 scales for interpreting group-level changes over time within a group of glioma patients, and for interpreting the mean difference in change between two groups, using both anchor-based and distribution-based approaches. We found that MIDs for the EORTC QLQ-C30 scales generally ranged between 4 and 11 points for both within-group mean change and between-group mean difference in change, representing small changes. The current estimates can be used to better interpret the results of clinical trials in glioma patients using the EORTC QLQ-C30 questionnaire.

Despite multimodal anti-cancer treatment, patients with primary and metastatic brain cancer have a poor prognosis. The most common malignant primary brain tumors are gliomas, with median survival rates ranging between 15 months and >15 years, depending on the histological tumor type, malignancy grade, and tumor genetics.[1] For these patients, the quality of survival is of utmost importance.

Health-related quality of life (HRQOL) is typically included as a secondary outcome in brain tumor clinical trials,[2–8] as a key component of determining the net clinical benefit of a treatment strategy. This means that the benefits of possible prolonged survival are weighed against the possible negative effects of treatment on the patients' level of functioning and well-being. The most common instruments used in brain tumor clinical trials to evaluate HRQOL are the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire Core 30 (QLQ-C30),[9] Functional Assessment of Cancer Therapy—General (FACT-G),[10] and the MD Anderson Symptom Inventory,[11] typically combined with their brain-specific modules.[12–14]

The impact of an experimental treatment on HRQOL is usually evaluated by comparing mean HRQOL scores over time in the experimental vs the control arm. Some studies rely on statistical significance alone to assess differences between treatment arms.[3,15] However, to demonstrate a true impact, differences should be both statistically significant and clinically relevant. For some questionnaires, minimally important differences (MIDs) have been determined, allowing to interpret differences and changes in HRQOL as clinically meaningful. Indeed, an MID can be considered as the smallest change/difference in a HRQOL score that is perceived as relevant or important by the patient and would indicate a change in the patient's disease management. Historically, for the EORTC QLQ-C30 questionnaire, a crude MID between 5 and 10 points was suggested to represent a small difference, between 10 and 20 points a moderate difference, and >20 a large difference, for both mean changes over time[16] and for mean differences between groups.[17] Therefore, in earlier studies, a difference of >10 points was considered clinically relevant for all scales. More recent work, however, suggests that these MIDs are too simplistic;[18,19] they do not distinguish between domain scales, direction of change (ie, improvement or deterioration), and disease sites. Therefore, MIDs for each EORTC QLQ-C30 scale specifically for glioma patients are needed, as they will be helpful in evaluating whether observed differences in HRQOL scores over time between treatment arms in a clinical trial are also clinically meaningful.

There are two approaches to estimating MIDs. The anchor-based approach links HRQOL measures to external criteria, either a factor that has clinical relevance (eg, performance status [PS]) or patient- or physician-reported ratings of health (eg, adverse events). The distribution-based approach, on the other hand, relies on the statistical features of HRQOL data (eg, standard deviation [SD]), and is typically used as supportive evidence for anchor-based estimates.[20] Maringwa et al. previously published MIDs for brain tumor patients based on two anchors, PS and Mini-Mental State Examination (MMSE) scores.[21] However, these anchors were suitable for only seven selected HRQOL scales. MID estimates for the selected scales, derived for both improvement and deterioration, were found to range between 5 and 14 points,[21] providing further evidence that the 10-point threshold is overly simplistic.

The aim of this study was to estimate the MIDs for all EORTC QLQ-C30 scales for interpreting group-level changes over time within a group of glioma patients, and for interpreting the mean difference in change between two groups, using both anchor-based and distribution-based approaches.

Materials and Methods

Data Description

Data were derived from three published EORTC phase III trials in glioma patients. Trial 1 (EORTC 26053) evaluated the impact of concurrent and/or adjuvant temozolomide chemotherapy to radiotherapy on overall survival in newly diagnosed non-1p/19q deleted anaplastic gliomas (n = 745).[22] Trial 2 (EORTC 26951) evaluated whether adjuvant procarbazine, lomustine, and vincristine chemotherapy improved overall survival compared to radiotherapy alone in newly diagnosed patients with anaplastic oligodendrogliomas or anaplastic oligoastrocytomas (n = 368).[23] Trial 3 (EORTC 26981) evaluated whether radiotherapy with concomitant and adjuvant temozolomide improved overall survival compared to radiotherapy alone in newly diagnosed glioblastoma patients (n = 573).[24] HRQOL was longitudinally assessed in all trials using the EORTC QLQ-C30 questionnaire. All trials were approved by the ethical committee of each participating center and all patients gave their informed consent to participate in the respective study.

EORTC QLQ-C30

The EORTC QLQ-C30 questionnaire[9] contains 30 items, 24 of which are combined into nine multi-item scales, ie, five functional scales (physical functioning [PF], role functioning [RF], cognitive functioning [CF], emotional functioning [EF], and social functioning [SF]) and three symptom scales (fatigue [FA], pain [PA], and nausea/vomiting [NV]), and one global health status (QL) scale. The remaining six single-item scales assess symptoms: dyspnea (DY), appetite loss (AP), insomnia (SL), constipation (CO), diarrhea (DI), and financial difficulties (FI). Trials 1 and 3 used version 3 of the EORTC QLQ-C30, whereas trial 2 used version 2. The two versions differ only in the response categories of questions 1-5 (ie, the PF scale), which were yes/no in version 2, whereas version 3 uses a four-point Likert scale ranging from "not at all" to "very much." The scoring of the EORTC QLQ-C30 scales was done according to the EORTC QLQ-C30 scoring manual,[25] with the means of the raw scores for each scale transformed to fall between 0 and 100. To facilitate interpretation, all scales were scored such that 0 represents the worst possible score and 100 the best possible score. The FI scale was omitted from the analysis because appropriate anchors were not obtainable, leaving 14 scales for analyses.

Clinical Anchors

Anchors for each EORTC QLQ-C30 scale were selected from clinical variables available in the datasets from the three source trials, including Common Terminology Criteria for Adverse Events (CTCAE), WHO PS, and MMSE, based on clinical plausibility and cross-sectional correlation.[20]

First, we assessed correlation. Depending on the distribution of the HRQOL scale/anchor pair, either a polyserial or polychoric correlation was estimated. Anchors with correlations of ≥|0.30|[26] were prioritized and where achievable, anchors with much stronger correlations were targeted.[27] The retained anchors were then assessed by a panel of five experts in both brain cancer and HRQoL for clinical plausibility.[28]

Definition of Clinical Change Groups

For each scale and time period, patients were categorized into one of the three clinical change groups: (i) deterioration (worsened by one anchor category), (ii) stable (no change in anchor categories), and (iii) improvement (improved by one anchor category). Patients who changed by ≥2 categories of an anchor were excluded because they were considered to be above the "minimal" expected change.

Data Analysis

Anchor-based method

The data were pooled across all three trials. Next, change scores for each EORTC QLQ-C30 scale and anchor pair were computed across all available pairwise time points during the entire follow-up period, and then combined into one dataset to provide sufficient data for examining clinically relevant changes. As an example, for a patient measured at time points t_a, t_b, and t_c, change scores were computed between t_a & t_b, t_a & t_c, and t_b & t_c. This meant that one patient could contribute to multiple change scores, and given their change scores, patients could contribute to multiple clinical change groups.

The mean change method was used to estimate MIDs for interpreting change over time *within* a group of patients. MIDs for improvement and deterioration corresponded to the mean HRQOL change scores within the improvement and deterioration clinical change groups, respectively. Moreover, for a given HRQOL scale, an effect size (ES) was computed within each clinical change group by dividing the mean of the HRQOL change scores (derived from all the pairwise time point differences) by the SD of the HRQOL change scores over all time points. Cohen[29] recommended that an ES of 0.2 is small, 0.5 is moderate, and ≥0.8 is large. Based on this guideline, only mean changes with an ES ≥0.2 and <0.8 were considered appropriate for inclusion as MIDs. The rationale was that an observed ES <0.2 reflects changes that are clinically unimportant, while ESs ≥0.8 were considered to reflect changes more than minimally important.

A linear regression approach was used to estimate MIDs for interpreting differences in change score over time *between* two distinct groups of patients. For a given EORTC QLQ-C30 scale, the outcome variable was the scale change score, and the covariate was a binary anchor variable with categories "stable" = 0 and "improvement" = 1 when modeling improvement (deteriorated observations were excluded), and a similar procedure was adopted for

deterioration. Since some patients contributed change scores to multiple clinical change groups, and more than one change score to a particular clinical change group, we corrected for the association between multiple change scores contributed by these patients by specifying a suitable covariance structure using generalized estimating equations (GEE).[30] The MIDs for improvement and deterioration correspond to the slope parameters for the "improved" and "deteriorated" covariates, respectively, and are useful for interpreting differences in change scores between groups of patients. Multiple MID values for a given scale were summarized to single values via correlation-based weighted average. To check whether using data from different trials had an impact on the estimated MIDs, this trial effect (ie, interaction between the trial and the anchor indicator) was included as covariate in the regression models. To account for multiple testing across the EORTC QLQ-C30 scales, we chose a priori to apply more stringent P values, ie, P values below .001 were considered statistically significant.

Distribution-based method

Two distribution-based methods were used, with calculations for each QLQ-C30 scale based on cross-sectional analysis; ie, the time point before or on the first day of treatment administration (t_1). Three proportions of an SD (0.2 SD, 0.3 SD, 0.5 SD) were calculated, representing the range of proportions currently considered relevant to MID estimation,[18,31,32] although there is no consensus on which best approximates the MID. The standard error of measurement (SEM) was calculated as SD multiplied by square root of ($1–r$), using SD at t_1. The test-retest reliability estimates (r) for the QLQ-C30 scales were obtained from Hjermstad et al.[33]

All statistical analyses were performed using SAS version 9.4.[34] In addition, two practical examples are provided to explain how MID results can be used to interpret HRQOL data from clinical trials and how MIDs can be used for a sample size calculation for a clinical trial in which HRQOL is the primary endpoint.

Results

A total of 1687 patients were enrolled across the three trials. Baseline characteristics are summarized in Table 1. Supplementary Figure S1 presents an overview of patient inclusion for the anchor-based method.

Anchor-Based Method

The final list of retained anchors, summarized in Table 2, comprised PS, scored between 0 (no symptoms of cancer) and 4 (bedbound), and eight CTCAEs (pain, fatigue, nausea and vomiting, gastrointestinal, constipation, anorexia, dyspnea, and neurology), which are graded between 0 (no toxicity) and 4 (life-threatening). At least one clinical anchor could be retained for 12/14 EORTC QLQ-C30 scales assessed, with insomnia and diarrhea therefore being excluded from further analysis.

Table 1 Baseline Patient Sociodemographic and Clinical Characteristics for Each Trial Separately and for the Total Population

	Trial			Total (N = 1697)
	26053 (N = 756)	26951 (N = 368)	26981 (N = 573)	
	N (%)	N (%)	N (%)	N (%)
Gender				
Male	448 (59.3)	212 (57.6)	360 (62.8)	1020 (60.1)
Female	308 (40.7)	156 (42.4)	213 (37.2)	677 (39.9)
WHO performance status				
0	440 (58.2)	138 (37.5)	223 (38.9)	801 (47.2)
1	284 (37.6)	172 (46.7)	277 (48.3)	733 (43.2)
2	32 (4.2)	58 (15.8)	73 (12.7)	163 (9.6)
Region				
Northern Europe	0 (0.0)	15 (4.1)	1 (0.2)	16 (0.9)
Southern Europe	64 (8.4)	43 (11.7)	57 (9.9)	164 (9.7)
Western Europe	482 (63.8)	300 (38.3)	324 (56.5)	1106 (65.2)
Eastern Europe	0 (0.0)	10 (2.7)	6 (1.0)	16 (0.9)
Non-European	210 (27.8)	0 (0.0)	185 (32.3)	395 (23.3)
Age				
Mean (SD)	43.01 (13.17)	47.62 (11.11)	53.81 (10.39)	47.65 (12.76)
Interquartile	32.0-52.0	40.0-55.0	48.0-62.0	37.0-58.0

Northern Europe: Sweden, Finland; Southern Europe: Italy, Spain; Western Europe: the Netherlands, France, United Kingdom, Germany, Belgium, Switzerland, Austria; Eastern Europe: Hungary, Poland, and Slovenia; Non-European: Canada, United States, Australia, Israel, and Turkey.

Table 2 Cross-Sectional Correlations of the EORTC QLQ-C30 Scales With Anchors and Correlations Between Their Change Scores

Scale	Anchor	Cross-Sectional		Change Scores	
		N (N$_o$)	Correlation	N (N$_o$)	Correlation
Physical functioning	Performance status	1317 (9687)	−0.52	921 (66765)	−0.31
Role functioning	Performance status	1325 (9655)	−0.51	921 (66551)	−0.24
Social functioning	Performance status	1330 (9660)	−0.41	921 (66209)	−0.20
Emotional functioning	Performance status	1326 (9669)	−0.30	921 (66288)	−0.20
Cognitive functioning	Performance status	1331 (9682)	−0.39	921 (66315)	−0.20
Cognitive functioning	CTCAE neurological	653 (3803)	−0.30	548 (18407)	−0.20
Global health status	Performance status	1330 (9653)	−0.43	921 (66045)	−0.20
Global health status	CTCAE fatigue	1179 (5577)	−0.33	975 (21129)	−0.20
Pain	CTCAE pain	654 (4015)	−0.44	548 (18570)	−0.30
Fatigue	Performance status	1334 (9701)	−0.44	975 (66662)	−0.30
Fatigue	CTCAE fatigue	1178 (5616)	−0.43	975 (21393)	−0.30
Nausea and vomiting	CTCAE nausea vomiting	1182 (5630)	−0.55	975 (21431)	−0.38
Nausea and vomiting	CTCAE gastrointestinal	1182 (5630)	−0.42	677 (21431)	−0.30
Appetite loss	CTCAE nausea vomiting	1182 (5616)	−0.33	975 (21359)	−0.20
Appetite loss	CTCAE gastrointestinal	1182 (5616)	−0.36	677 (21359)	−0.22
Appetite loss	CTCAE anorexia	1181 (5614)	−0.52	975 (21355)	−0.33
Dyspnea	CTCAE dyspnea	654 (3812)	−0.44	548 (18511)	−0.21
Constipation	CTCAE gastrointestinal	1179 (5597)	−0.30	677 (21232)	−0.20
Constipation	CTCAE constipation	526 (1795)	−0.53	427 (2800)	−0.40

Abbreviations: CTCAE, Common Terminology Criteria for Adverse Events; EORTC QLQ-C30, European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30.
N = number of patients, N$_o$ = number of HRQOL observations.

The cross-sectional correlations between the different HRQOL scales and anchors ranged from 0.30 to 0.55 in absolute value, and the correlations between their change scores ranged from 0.20 to 0.40 (Table 2). Supplementary Table S1 provides the distribution of patients and the number of change observations across change categories of the various anchors. According to the anchor change scores, most patients were stable on all anchors. The number of patients who either improved or deteriorated was similar for most anchors except for PS, in which more patients deteriorated than improved. Anchor-based MIDs with clinically important ES could be determined for deterioration in 12 EORTC QLQ-C30 scales assessed, and in 11 scales for improvement. The anchor-based MIDs for change within-group and between-group differences in change over time are presented in Table 3. Supplementary Table S2 reports full details for both the mean change method (within-group change) and the linear regression method (between-group difference in change). Note that only MIDs estimated from anchor clinical change groups with a clinically important ES ≥0.2 and <0.8 contributed to MID estimates in Table 3.

Figure 1 presents MIDs from the mean change method (see also Table 3), plotted alongside their 95% confidence intervals, showing how MIDs vary by EORTC QLQ-C30 scale, direction of change (improvement vs deterioration), and anchor. All MID estimates were always in the expected direction, ie, positive vs negative mean change scores within the improvement vs deterioration clinical change group respectively. MIDs for interpreting within-group change (derived from the mean change method) ranged from 4 to 12 points for improvement and −14 to −4 points for deterioration. MIDs for between-group difference in change (derived from the linear regression) ranged from 4 to 9 for improvement and −16 to −4 deterioration (Table 3).

Taking into account the weighted summaries, MIDs for the EORTC QLQ-C30 scales generally ranged from 4 to 11 points in absolute values for both within-group and between-group changes. Adding the trial effect to the regression models showed no statistically significant differences for improving or deteriorating scores (data not shown), supporting the combination of the three trials.

Distribution-Based Method

Distribution-based estimates for all 14 EORTC QLQ-C30 scales considered in our analyses are presented in Supplementary Table S3. Most anchor-based MIDs were closest to the distribution-based estimates when using 0.3 SD as cutoff, ranging between 3.40 and 9.86.

Practical Examples

Table 4 presents two practical example examples on how MIDs can be used to interpret HRQOL results in a glioma clinical trial (example 1), or to calculate the sample size for

Table 3   Summary of Anchor-Based MIDs (Weighted Average) for Within-Group and Between-Group Difference in Change Over Time

Scale	Within-Group Change[a]		Between-Group Difference in Change[a]	
	Improvement	Deterioration	Improvement	Deterioration
Physical functioning	5	−9	5	−7
Role functioning	9	−9	8	−9
Social functioning	6	−6	5	−6
Emotional functioning	6	−5	4	−4
Cognitive functioning	No MID	−9 to −5 (−7)	No MID	−6.1 to −5.0 (−5.5)
Global health status	4 to 6 (5)[b]	−6	3.6 to 5.3 (4.4)	−6
Fatigue	8.2 to 9.2 (8.7)[b]	−8 to −6 (−7)	7.5 to 7.7 (7.6)	−8 to −7 (−7.4)
Pain	6	−8	7	−6
Nausea and vomiting	7	−7	6.42	−7
Dyspnea	9	−8	6.64	−8
Appetite loss	10.9 to 11.7 (11.3)[b]	−5.3 to −4.3 (−4.8)	9.21 to 9.24 (9.22)	−7.6 to −7.4 (−7.5)
Constipation	5	−14 to −9 (−10)	5	−16 to −7 (−10)

Abbreviation: MIDs, minimally important differences.

[a]The within-group MIDs are derived from the mean change method and the between-group MIDs from the linear regression.

[b]The numbers between brackets represent the weighted correlations, and only applies to scales with more than one anchor retained.

Note: The symptom scores were reversed to follow the functioning scales' interpretation, ie, 0 represents the worst possible score and 100, the best possible score; "no MID" is used where no MID estimate is available either due to the absence of a suitable anchor or effect size <0.2 or ≥0.8.

a clinical trial in glioma patients in which HRQOL is the primary endpoint (example 2).

Discussion

The MIDs for most scales of the EORTC QLQ-C30 estimated in this study can be used to interpret group-level changes over time in glioma patients. MIDs generally ranged from 4 to 11 points, both for change within-group and between-group differences in change. Moreover, the magnitude of the MIDs per scale was fairly similar for both deterioration and improvement.

The results of this study underline the assumption that using one MID as clinically relevant for all EORTC QLQ-C30 scales in all cancers is too simplistic. Although the MIDs for the direction of change were fairly similar for each scale, there was a wide range in MIDs between the different scales in this brain tumor population. For only a minority of scales, the MID was >10 points,[16] meaning that applying this cutoff may possibly have resulted in an underestimation of clinically relevant differences between and within groups in previously analyzed brain cancer trials. Our results also highlight that MIDs differ across disease sites. For example, QLQ-C30 MIDs for (advanced) cancer patients were not equal to those of glioma patients.[26,35,36]

It should be emphasized that the presented MIDs are developed for interpreting group-level results of analyses based on HRQOL scales as continuous variables. They may not be applicable to analyses of HRQOL-derived metrics at the individual patient level (ie, responder analyses), eg, reporting the proportion of patients that changed in HRQOL by a clinically important degree in a specific time period or the time to HRQOL deterioration. A project to establish

so-called responder thresholds for changes in HRQOL on the individual patient level is currently ongoing.[28]

Previously, Maringwa et al.[21] estimated MIDs for five scales of the EORTC QLQ-C30 and two scales of the brain cancer-specific EORTC QLQ-BN20 using two trials that are also used in this study. They found MIDs ranging between 5 and 14 points. For the scales that could be compared between studies, the MIDs found by Maringwa et al. were generally a few points higher than those found in our study. Differences may be explained by the number and type of anchors that were used, as well as methodological differences, ie, number of time points to calculate changes and number of patients within each clinical change group. In contrast to the study of Maringwa et al., the current study included all scales of the EORTC QLQ-C30, except for financial difficulties, insomnia or diarrhea for which no anchor was available or could be established, and MIDs for both change within-group and between-group differences in change were calculated. In the current study, brain tumor-specific functioning and symptoms as measured with the EORTC QLQ-BN20 were not included, while these may be particularly relevant for brain tumor patients. Currently, the EORTC QLQ-BN20 is undergoing a revision, as this questionnaire has been developed in 1996[12] with an international validation in 2010,[13] and treatments have changed since then. These new treatments bring new toxicities, eg, eye problems,[37] which are not sufficiently covered by the current version. For the revised brain cancer module, it would also be important to determine scale-specific MIDs that can be used to interpret differences or changes in HRQOL scores in brain cancer clinical trials. It should be noted though, that the clinical trial data that were used in this study was based on glioma patients only. The MIDs derived in this study may therefore not be applicable to all brain cancer populations, including those
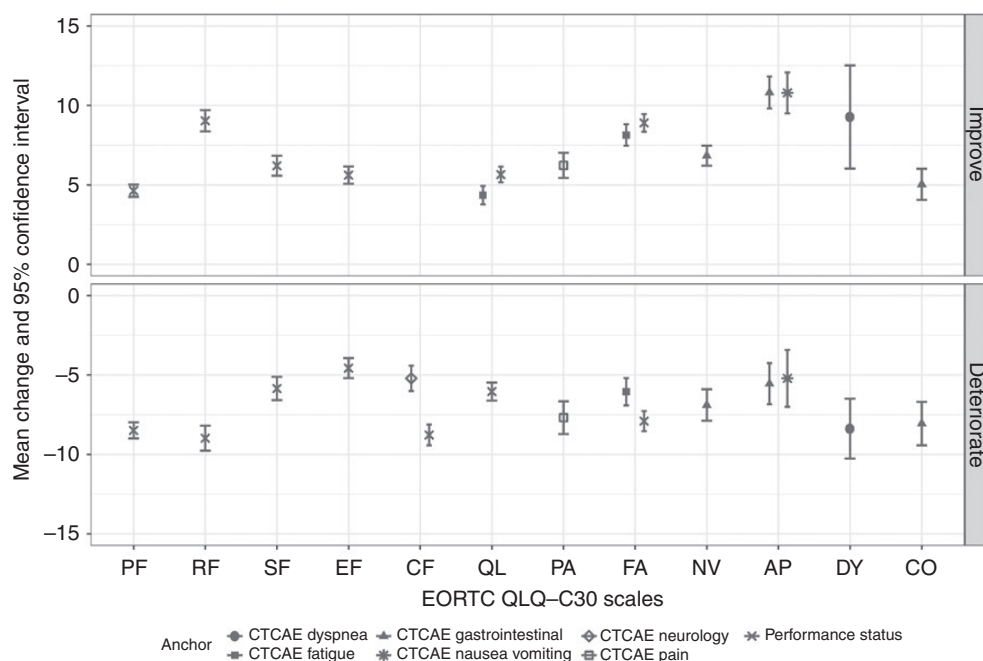
Fig. 1 Mean change and 95% confidence interval for improvement and deterioration in EORTC QLQ-C30 scales, across multiple anchors. Estimates are available only for scales with at least 1 suitable anchor and with effect size >=0.2 and <0.8. These mean change scores are useful for interpreting within-group change over time. Abbreviation: AP, appetite loss; CO, constipation; CF, cognitive functioning; CTCAE, common terminology criteria for adverse events; DY, dyspnea; EF, emotional functioning; FA, fatigue; NV, nausea/vomiting; PA, pain; PF, physical functioning, QL, global health status; RF, role functioning; SF, social functioning. Deteriorate = worsened by at least 1 anchor category; no change = no change in anchor categories; improve = improved by at least 1 category.

Table 4 Practical Examples on How Minimally Important Differences (MIDs) Can Be Used to (1) Interpret Health-Related Quality of Life (HRQOL) Results in a Glioma Clinical Trial or (2) to Calculate the Sample Size for a Clinical Trial in Glioma Patients in Which HRQOL Is the Primary Endpoint

Practical example 1

In this example, we will illustrate how the MIDs can be used to interpret HRQOL results in a clinical trial. In a clinical trial comparing treatment A with treatment B, a statistically significant mean difference in change of 3 points between the two treatment arms was found for the physical functioning scale (both arms deteriorated in their level of physical functioning, treatment A with 2 points and treatment B with 5 points). The estimated MID for deterioration of physical functioning was found to be 7 points (Table 3). This means that although the observed difference of 3 points between treatment arms was statistically significant, this difference cannot be considered clinically meaningful and would not require a change in patient management. In this case, we should conclude that the treatments were similar in their impact on the patients' HRQOL. Such a scenario is quite plausible in a trial with a large sample size, where HRQOL comparisons will be statistically overpowered.

Practical example 2

Another application of the estimated MIDs is their use in the sample size calculation for a clinical trial in glioma patients in which HRQOL is the primary endpoint. For example, it is expected that the deterioration in physical functioning with treatment A will be less pronounced than the deterioration in physical functioning with treatment B. The difference in deterioration between the treatment arms is expected to be clinically meaningful. In that case, we could use the MID for deterioration, which is 7 points, as the mean difference that is needed to detect a clinically meaningful difference between the treatment arms, and power/calculate the sample size accordingly.

with metastatic brain cancer, in which the EORTC QLQ-C30 is often used. Also, the EORTC questionnaires are regularly used in studies including patients with benign brain tumors, such as meningioma.[38–40] Whether these MIDs are also applicable to other types of brain tumors therefore remains to be determined. Another limitation in clinical trials that may have impacted the estimations of the MIDs is missing HRQOL data. In all three trials, compliance rates with HRQOL assessment decreased over time, which

is likely related to the health status of the patient. Patients with a better health status may have been overrepresented during follow-up, hampering generalizability of our results.

Although the correlations between the HRQOL scales and anchors were not strong, ie, between 0.2 and 0.4 for the change scores, estimates from the distribution-based method using the 0.3 SD cutoff were comparable to those derived with the anchor-based method. Currently, there is no consensus on which SD best approximates the MID,[18,31,32]

but 0.2 SD has previously been suggested to reflect a clinically important small effect,[29] and 0.3 SD has been suggested to reflect an appropriate threshold for defining MIDs.[41] Also, the estimated MIDs were often within a relatively small range and in the expected direction of change similar to the anchor, and the final MID estimates were based on correlation weighted averages thereby mitigating the influence of weaker anchors, further supporting the use of the chosen anchors. Moreover, the anchors that were retained based on statistical considerations were subsequently reviewed by experts on their clinical plausibility to avoid spurious findings. However, not all (clinically) appropriate anchors may have been available to determine MIDs, since we used existing data of closed clinical trials. For example, we were not able to establish an MID for improvements in the cognitive functioning scale, while this is an important outcome for brain tumor patients, as over 80% of patients with primary or metastatic brain cancer have neurocognitive impairments at diagnosis.[42–44] In this study, PS and CTCAE neurological were used as clinical anchors for cognitive functioning, but change scores in cognitive functioning had weak correlations with change in the respective anchors, both –0.20. This may be due to the fact that both anchors were physician-reported, which are subjective by nature. Although objective neurocognitive functioning is often measured in brain tumor trials, the correlation between objectively measured neurocognitive functioning and cognitive complaints as measured with a patient-reported outcome is generally poor.[45–47] This was also true for this study; MMSE was considered as a clinical anchor as this outcome was measured in all trials, but the correlations were <0.30 and therefore excluded from further analyses. Using a patient-reported outcome to assess cognitive complaints, eg, MOS (Medical Outcomes Study) Cognitive Functioning Scale,[48] may have been a more appropriate anchor. Otherwise, more objective anchors should be identified and used in clinical trials.

In conclusion, our study provides estimates for MIDs for group-level interpretation of the EORTC QLQ-C30 in glioma patients. In general, the estimates range between 4 and 11 points and correspond to small differences as proposed by Cocks et al. in the guidelines for interpretation of longitudinal HRQOL differences.[18] The current estimates can be used to better interpret results of clinical trials in glioma patients, and subsequently inform patients, as well as for sample size calculations for planning future glioma clinical trials. Advances to establish MIDs to evaluate changes in HRQOL on the individual patient level, which is relevant in clinical practice, are currently ongoing.[49]

Supplementary Material

Supplementary material is available at *Neuro-Oncology* online.

Keywords

brain tumor | clinical relevance | EORTC QLQ-C30 | health-related quality of life (HRQOL) | minimally important difference (MID)

References

1. Ostrom QT, Gittleman H, Fulop J, et al. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2008–2012. *Neuro Oncol.* 2015;17(Suppl 4): iv1–iv62.

2. Armstrong TS, Wefel JS, Wang M, et al. Net clinical benefit analysis of radiation therapy oncology group 0525: a phase III trial comparing conventional adjuvant temozolomide with dose-intensive temozolomide in patients with newly diagnosed glioblastoma. *J Clin Oncol.* 2013;31(32):4076–4084.

3. Keime-Guibert F, Chinot O, Taillandier L, et al.; Association of French-Speaking Neuro-Oncologists. Radiotherapy for glioblastoma in the elderly. *N Engl J Med.* 2007;356(15):1527–1535.

4. Minniti G, Scaringi C, Baldoni A, et al. Health-related quality of life in elderly patients with newly diagnosed glioblastoma treated with short-course radiation therapy plus concomitant and adjuvant temozolomide. *Int J Radiat Oncol Biol Phys.* 2013;86(2):285–291.

5. Taphoorn MJ, Henriksson R, Bottomley A, et al. Health-related quality of life in a randomized phase III study of bevacizumab, temozolomide, and radiotherapy in newly diagnosed glioblastoma. *J Clin Oncol.* 2015;33(19):2166–2175.

6. Taphoorn MJ, Stupp R, Coens C, et al.; European Organisation for Research and Treatment of Cancer Brain Tumour Group; EORTC Radiotherapy Group; National Cancer Institute of Canada Clinical Trials Group. Health-related quality of life in patients with glioblastoma: a randomised controlled trial. *Lancet Oncol.* 2005;6(12):937–944.

7. Taphoorn MJ, van den Bent MJ, Mauer ME, et al.; European Organisation for Research and Treatment of Cancer. Health-related quality of life in

patients treated for anaplastic oligodendroglioma with adjuvant chemotherapy: results of a European Organisation for Research and Treatment of Cancer randomized clinical trial. *J Clin Oncol.* 2007;25(36):5723–5730.

8. Taphoorn MJB, Dirven L, Kanner AA, et al. Influence of treatment with tumor-treating fields on health-related quality of life of patients with newly diagnosed glioblastoma: a secondary analysis of a randomized clinical trial. *JAMA Oncol.* 2018;4(4):495–504.

9. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst.* 1993;85(5):365–376.

10. Cella DF, Tulsky DS, Gray G, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *J Clin Oncol.* 1993;11(3):570–579.

11. Armstrong TS, Mendoza T, Gning I, et al. Validation of the M.D. Anderson Symptom Inventory Brain Tumor Module (MDASI-BT). *J Neurooncol.* 2006;80(1):27–35.

12. 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(1):139–150.

13. Taphoorn MJ, Claassens L, Aaronson NK, et al.; EORTC Quality of Life Group, and Brain Cancer, NCIC and Radiotherapy Groups. An international validation study of the EORTC brain cancer module (EORTC QLQ-BN20) for assessing health-related quality of life and symptoms in brain cancer patients. *Eur J Cancer.* 2010;46(6):1033–1040.

14. Weitzner MA, Meyers CA, Gelke CK, Byrne KS, Cella DF, Levin VA. The Functional Assessment of Cancer Therapy (FACT) scale. Development of a brain subscale and revalidation of the general version (FACT-G) in patients with primary brain tumors. *Cancer.* 1995;75(5):1151–1161.

15. Shields LB, Choucair A, Choucair AK. Quality of life measures as a preliminary clinical indicator in patients with primary brain tumors. *Surg Neurol Int.* 2013;4:48.

16. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol.* 1998;16(1):139–144.

17. King MT. The interpretation of scores from the EORTC Quality of Life Questionnaire QLQ-C30. *Qual Life Res.* 1996;5(6):555–567.

18. Cocks K, King MT, Velikova G, et al. Evidence-based guidelines for interpreting change scores for the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. *Eur J Cancer.* 2012;48(11):1713–1721.

19. Cocks K, King MT, Velikova G, Martyn St-James M, Fayers PM, Brown JM. Evidence-based guidelines for determination of sample size and interpretation of the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. *J Clin Oncol.* 2011;29(1):89–96.

20. 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(2):102–109.

21. Maringwa J, Quinten C, King M, et al.; EORTC PROBE Project and Brain Cancer Group. Minimal clinically meaningful differences for the EORTC QLQ-C30 and EORTC QLQ-BN20 scales in brain cancer patients. *Ann Oncol.* 2011;22(9):2107–2112.

22. van den Bent MJ, Baumert B, Erridge SC, et al. Interim results from the CATNON trial (EORTC study 26053-22054) of treatment with concurrent and adjuvant temozolomide for 1p/19q non-co-deleted anaplastic glioma: a phase 3, randomised, open-label intergroup study. *Lancet.* 2017;390(10103):1645–1653.

23. 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(18):2715–2722.

24. Stupp R, Mason WP, van den Bent MJ, et al.; European Organisation for Research and Treatment of Cancer Brain Tumor and Radiotherapy Groups; National Cancer Institute of Canada Clinical Trials Group. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med.* 2005;352(10):987–996.

25. Fayers PM, Aaronson NK, Bjordal K, Curran D, Groenvold M, Bottomley A, on behalf of the EORTC Quality of Life Group. *EORTC QLQ-C30 Scoring Manual.* 3rd ed. Brussels: EORTC Quality of Life Group; 2001.

26. Musoro JZ, Bottomley A, Coens C, et al.; EORTC Melanoma Group and EORTC Quality of Life Group. Interpreting European Organisation for Research and Treatment for Cancer Quality of Life Questionnaire Core 30 scores as minimally importantly different for patients with malignant melanoma. *Eur J Cancer.* 2018;104:169–181.

27. Coon CD. Empirical telling the interpretation story: the case for strong anchors and multiple methods. 23rd Annual Conference of the International Society for Quality of Life Research, Copenhagen, Denmark. *Qual Life Res.* 2016;25:1–2.

28. Musoro ZJ, Hamel JF, Ediebah DE, et al.; EORTC Quality of Life Group. Establishing anchor-based minimally important differences (MID) with the EORTC quality-of-life measures: a meta-analysis protocol. *BMJ Open.* 2018;8(1):e019117.

29. Cohen J. *Statistical Power Analysis for the Behavioural Sciences.* 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

30. Liang KY, Zeger SL. Regression analysis for correlated data. *Annu Rev Public Health.* 1993;14:43–68.

31. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res.* 2011;11(2):171–184.

32. 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(5):582–592.

33. 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(5):1249–1254.

34. SAS Institute Inc. *Base SAS® 9.4 Procedures Guide.* Second Edition. Cary, NC: SAS Institute Inc. 2013.

35. Bedard G, Zeng L, Zhang L, et al. Minimal important differences in the EORTC QLQ-C30 in patients with advanced cancer. *Asia Pac J Clin Oncol.* 2014;10(2):109–117.

36. Musoro JZ, Coens C, Fiteni F, et al.; EORTC Breast and Quality of Life Groups. Minimally important differences for interpreting EORTC QLQ-C30 scores in patients with advanced breast cancer. *JNCI Cancer Spectr.* 2019;3(3):pkz037.

37. van den Bent M, Gan HK, Lassman AB, et al. Efficacy of depatuxizumab mafodotin (ABT-414) monotherapy in patients with EGFR-amplified, recurrent glioblastoma: results from a multi-center, international study. *Cancer Chemother Pharmacol.* 2017;80(6):1209–1217.

38. Erharter A, Giesinger J, Kemmler G, et al. Implementation of computer-based quality-of-life monitoring in brain tumor outpatients in routine clinical practice. *J Pain Symptom Manage.* 2010;39(2):219–229.

39. Konglund A, Rogne SG, Lund-Johansen M, Scheie D, Helseth E, Meling TR. Outcome following surgery for intracranial meningiomas in the aging. *Acta Neurol Scand.* 2013;127(3):161–169.

40. Nassiri F, Price B, Shehab A, et al.; International Consortium on Meningiomas. Life after surgical resection of a meningioma: a prospective cross-sectional study evaluating health-related quality of life. *Neuro Oncol.* 2019;21(Suppl 1):i32–i43.

41. 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(9):898–910.

42. Meyers CA, Smith JA, Bezjak A, et al. Neurocognitive function and progression in patients with brain metastases treated with whole-brain radiation and motexafin gadolinium: results of a randomized phase III trial. *J Clin Oncol.* 2004;22(1):157–165.

43. Tucha O, Smely C, Preier M, Lange KW. Cognitive deficits before treatment among patients with brain tumors. *Neurosurgery.* 2000;47(2):324–334; discussion 333.

44. van Loenen IS, Rijnen SJM, Bruijn J, Rutten GM, Gehring K, Sitskoorn MM. Group changes in cognitive performance after surgery mask changes in individual patients with glioblastoma. *World Neurosurg.* 2018;117:e172–e179.

45. Gehring K, Taphoorn MJ, Sitskoorn MM, Aaronson NK. Predictors of subjective versus objective cognitive functioning in patients with stable grades II and III glioma. *Neurooncol Pract.* 2015;2(1):20–31.

46. Pranckeviciene A, Deltuva VP, Tamasauskas A, Bunevicius A. Association between psychological distress, subjective cognitive complaints and objective neuropsychological functioning in brain tumor patients. *Clin Neurol Neurosurg.* 2017;163:18–23.

47. van Lonkhuizen PJC, Rijnen SJM, van der Linden SD, Rutten GM, Gehring K, Sitskoorn MM. Subjective cognitive functioning in patients with a meningioma: its course and association with objective cognitive functioning and psychological symptoms. *Psychooncology.* 2019;28(8):1654–1662.

48. Hays RD, Sherbourne CD, Mazel R. *User's Manual for the Medical Outcomes Study (MOS) Core Measures of Health-Related Quality of Life.* Santa Monica, CA: RAND Corporation; 1995.

49. Giesinger JM, Loth FLC, Aaronson NK, et al.; EORTC Quality of Life Group. Thresholds for clinical importance were established to improve interpretation of the EORTC QLQ-C30 in clinical practice and research. *J Clin Epidemiol.* 2020;118:1–8.