

Minimal important differences for interpreting health-related quality of life scores from the EORTC QLQ-C30 in lung cancer patients participating in randomized controlled trials

John T. Maringwa · Chantal Quinten ·
Madeleine King · Jolie Ringash · David Osoba ·
Corneel Coens · Francesca Martinelli ·
Jurgen Vercauteren · Charles S. Cleeland ·
Henning Flechtner · Carolyn Gotay · Eva Greimel ·
Martin J. Taphoorn · Bryce B. Reeve ·
Joseph Schmucker-Von Koch · Joachim Weis ·
Egbert F. Smit · Jan P. van Meerbeeck ·
Andrew Bottomley ·
on behalf of the EORTC PROBE project and the Lung
Cancer Group

Received: 10 May 2010 / Accepted: 20 September 2010 / Published online: 1 October 2010
© Springer-Verlag 2010

Abstract

Background The aim of this study was to determine the smallest changes in health-related quality of life (HRQOL) scores in a subset of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire core 30 (EORTC QLQ-C30) scales, which could be considered as clinically meaningful in patients with non-small-cell lung cancer (NSCLC).

Methods WHO performance status (PS) and weight change were used as clinical anchors to determine minimal important differences (MIDs) in HRQOL change scores (range, 0–100) in the EORTC QLQ-C30 scales. Selected distribution-based methods were used for comparison.

Findings In a pooled dataset of 812 NSCLC patients undergoing treatment, the values determined to represent the MID depended on whether patients were improving or

J. T. Maringwa (✉) · C. Quinten · C. Coens · F. Martinelli ·
J. Vercauteren · A. Bottomley
Quality of Life Department, EORTC,
Brussels, Belgium
e-mail: john.maringwa@eortc.be

M. King
Psycho-oncology Co-operative Research Group,
University of Sydney,
Sydney, Australia

J. Ringash
The Princess Margaret Hospital, University of Toronto,
Toronto, Canada

D. Osoba
Quality of Life Consulting,
West Vancouver,
Vancouver, Canada

C. S. Cleeland
Department of Symptom Research, University of Texas,
Houston, TX, USA

H. Flechtner
Child and Adolescent Psychiatry and Psychotherapy,
University of Magdeburg,
Magdeburg, Germany

C. Gotay
School of Population and Public Health,
University of British Columbia,
Vancouver, Canada

E. Greimel
Obstetrics and Gynecology,
Medical University Graz,
Graz, Austria

M. J. Taphoorn
VU Medical Center,
Amsterdam, Netherlands

B. B. Reeve
Division of Cancer Control and Population Science,
National Cancer Institute,
Bethesda, USA

deteriorating. MID estimates for improvement (based on a one-category change in PS, 5–<20% weight gain) were physical functioning (9, 5); role functioning (14, 7); social functioning (5, 7); global health status (9, 4); fatigue (14, 5); and pain (16, 2). The respective MID estimates for deterioration (based on PS, weight loss) were physical (4, 6); role (5, 5); social (7, 9); global health status (4, 4); fatigue (6, 11); and pain (3, 7).

Interpretation Based on the selected QLQ-C30 scales, the MID may depend upon whether the patients' PS is improving or worsening, but our results are not definitive. The MID estimates for the specified scales can help clinicians and researchers evaluate the significance of changes in HRQOL and assess the value of a health care intervention or compare treatments. The estimates also can be useful in determining sample sizes in the design of future clinical trials.

Keywords Anchoring · EORTC QLQ-C30 · Health-related quality of life · Minimal important difference

Introduction

Determining the minimal important difference (MID) [1–4] for interpreting health-related quality of life (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 [5].

Methods aimed at identifying MIDs are classified as either anchor-based or distribution-based [6]. Anchor-based methods link HRQOL measures either to known indicators

that have clinical relevance (e.g., progression of disease, performance status (PS), etc.) or to patient-derived ratings of change in health [1, 6]. Distribution-based approaches hinge on summary statistics calculated from the HRQOL data; two commonly used statistics are the effect size [7] and standard error of measurement (SEM) [8]. The effect size used in the MID literature is the mean change divided by the between-person standard deviation; this aids interpretation by benchmarking the mean change against the degree of variation among individuals. An effect size of 0.2 standard deviations (SD) of HRQOL scores has been proposed as a definition for a minimal clinically important difference [9]. Some suggested that 0.5 SD is a reasonable approximation for the MID [10], although others feel this estimate is not generalizable [11, 12]. Thresholds of 1 SEM have also been used to estimate MIDs [13]. Other investigators, using data from patients' ratings of their own global change [14] or from patients' comparisons of themselves to others [15], determined that 5–10% of the instrument range represents a subjectively significant difference or clinically significant change.

It is important to determine the MID for various instruments (questionnaires) in a variety of cancers because a determination of *p* values does not provide information about the clinical meaningfulness of differences between groups or changes over time within a group. *P* values are highly dependent on sample size. In large sample sizes, significant *p* values can be obtained when numerical differences in HRQOL change scores are small and not likely to be clinically meaningful. As more and more studies examine the MIDs for differing questionnaires and cancers, it will become evident whether it is possible to generalize and adopt one MID or a set of MIDs for all questionnaires and patient groups. It will take a large number of such explorations to increase the confidence of investigators. Thus, every study contributing to this question is important.

The European Organization 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. Anchor-based methods have been used previously to inform the interpretation of QLQ-C30 scores [14, 16]. Using global ratings of change as the anchor, Osoba et al. [14] suggested that in patients with breast and small-cell lung cancers, 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 [16] obtained similar results when she collated 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

J. S.-V. Koch
Medical Ethics, University of Regensburg,
Regensburg, Germany

J. Weis
Tumor Biology Center, University Freiburg,
Freiburg, Germany

E. F. Smit
Department of Pulmonary Diseases, Vrije Universiteit VUMC,
Amsterdam, Netherlands

J. P. van Meerbeeck
Department of Respiratory Medicine, Ghent University Hospital,
Ghent, Belgium

M. J. Taphoorn
Medical Center Haaglanden,
The Hague, Netherlands

clinical trials that use the QLQ-C30 [17]. However, the evidence is not clear that a 10-point threshold is applicable to each of the 15 QLQ-C30 scales [17]. Further, it has not yet been established whether the same thresholds apply to improvement and deterioration in HRQOL scores. Additional empirical investigation of the size and patterns of MIDs across domains of the QLQ-C30 is therefore justified.

Our focus was to determine the change in selected QLQ-C30 scales which corresponds to the MID for improvement and deterioration in HRQOL for non-small-cell lung cancer (NSCLC) patients. Identification of MIDs was carried out using two clinical anchors: change in physician-rated WHO PS and weight change. Since no MIDs on the QLQ-C30 have been determined for NSCLC patients and since MIDs may vary across patient groups, we focused on NSCLC with the intention of analyzing other sites later.

Patients and methods

The EORTC QLQ-C30

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 but worse symptoms.

Description of the data and selection of QLQ-C30 scales

Two closed EORTC randomized controlled trials enrolling in total 812 palliative, locally advanced, and/or metastatic NSCLC patients were jointly analyzed. Trial 1 compared gemcitabine+cisplatin and paclitaxel+gemcitabine to the standard arm paclitaxel+cisplatin, enrolled 480 patients, and used the QLQ-C30 version 3 [18]. Trial 2 compared two cisplatin-based combination chemotherapies, enrolled 332 patients, and used QLQ-C30 version 1 [19]. These two versions of the QLQ-C30 differ only in the response options for the items in the physical and role functioning domains. Version 1 uses a binary (no, yes) scale and version 3 uses a four-point scale ranging from “not at all” to “very much” [20]. In both trials, HRQOL was measured as a secondary endpoint at baseline, during treatment, and on several follow-up occasions after the end of treatment.

Physical (PF), role (RF), and social (SF) functioning, global health status (GHS), fatigue (FA) and pain (PA) were

chosen for this analysis because they were expected to show relatively strong association with the chosen anchors. Indeed, correlations between the other scales and the anchors were relatively weak (data not shown).

Both trials involved the same cancer site and had similar treatment modalities; thus, data from these trials were pooled. Due to the mentioned differences in the versions of the QLQ-C30, analysis for PF and RF was restricted to trial 1 which used version 3, the current version [20].

Clinical anchors

The anchor-based approach to developing MIDs requires an anchor that is itself interpretable and at least moderately correlated with the instrument being explored [2]. The chosen clinical anchors are clearly definable and understandable, they are commonly used by clinicians in assessment of cancer patients, and they have previously been shown to be correlated with HRQOL assessments of cancer patients [16, 21, 22]. Values for the 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). Following CTCAE [23] guidelines, changes in weight were grouped as weight loss ($5 < 20\%$ loss), no change ($< 5\%$ loss or gain of total body weight), and weight gain ($5 < 20\%$ gain). Patients whose PS changed by two or more categories or body weight changed by $\geq 20\%$ (conventionally classified as severe loss [23]) were excluded since such changes were considered to be more than “minimal” in terms of their clinical relevance in this patient population.

Data analysis

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

Differences in the anchor values and HRQOL scores between T_1 and T_2 were calculated for each patient. Eleven patients who deteriorated by more than one PS category were excluded from the PS analysis. No patients improved by more than one PS category. Three patients who lost $\geq 20\%$ total body weight were excluded from the weight loss analysis. No patients gained $\geq 20\%$ of their weight.

The differences in individual patient’s HRQOL scores were then assigned to one of three “clinically meaningful” categories, as defined a priori by the anchors, e.g.,

“improvement”, “no change”, and “deterioration” groups for PS (see “Clinical anchors”). We obtained estimates of the MIDs by calculating the difference in mean HRQOL change between adjacent categories [13], i.e., “improvement” versus “no change” and “no change” versus “deterioration”. This was done to control for the amount of change in HRQOL that occurred to patients who did not change according to the anchor. The 95% confidence intervals (CI) for the differences in mean of change scores were calculated.

The association between HRQOL scores and anchor values, and between changes in both the anchor and HRQOL scale, was quantified by the Spearman rank correlation coefficient. Revicki et al. [24] suggested a correlation of at least 0.30 as a measure of an acceptable association.

For comparison purposes, three distribution-based approaches were applied: 0.5 SD, 0.20 SD, and the SEM. The SEM measures the precision of the HRQOL instrument [8]. We calculated SEM using SD at T₁ and T₂, separately, and test–retest reliability estimates provided by Hjørnstad et al. [25]. Our results were also compared with the 5–10% range of the instrument [15].

Results

Table 1 gives a summary of selected demographic and clinical characteristics of the patients at baseline in the combined data from the two trials.

Descriptive statistics summarizing the distributions of HRQOL scores at baseline are given in Table 2. The distributions for PF, SF, RF, PA, and FA were skewed, with a predominance of good functioning and low symptoms, while GHS was reasonably symmetrical. The mean and standard deviations of HRQOL scores at the two time points T₁ and T₂ are also given in Table 2.

From Table 2, the cross-sectional correlations of HRQOL measures with PS were generally moderate, ranging in absolute value from 0.30 to 0.44. Correlations for GHS and SF at T₁ (−0.29, −0.23) and PA at T₂ (0.24) were relatively weak. Except for appetite loss, cross-sectional correlations for all other scales of the QLQ-C30 with PS both at T₁ and T₂ were less than for the scales we chose (data not shown). For changes in HRQOL scores and changes in both anchors, the correlations were generally weak (ranging 0.03–0.21 in absolute value).

When anchoring with PS, the number of days between T₁ and T₂ ranged from 20 to 161 with a mean of 76 (SD=34.2) for trial 1. For trial 2, the number of days ranged from 20 to 194 with mean of 88 (SD=37.9). A very similar distribution for the time separation was observed when anchoring with weight change. Including the number of

Table 1 Selected baseline demographic and clinical characteristics of the patients

Descriptive statistics		Number	Percent
Gender	Male	552	68.0
	Female	260	32.0
Performance status	0	231	28.0
	1	491	61.0
	2	90	11.0
Country	Belgium	32	3.9
	Czech Republic	5	0.6
	Egypt	37	4.6
	France	5	0.6
	Germany	24	3.0
	Italy	50	6.2
	Poland	4	0.5
	South Africa	4	0.5
	Spain	74	9.1
	Switzerland	3	0.4
	The Netherlands	562	69.2
TNM staging	United Kingdom	12	1.5
	Stage IV	592	73.0
	Stage IIIB	183	22.0
Weight (n=780)	Other	37	5.0
	Mean (kg)		72
Age (n=812)	Interquartile range (kg)		63–80
	Mean (years)		57
	Interquartile range (years)		50–65

days between T₁ and T₂ as a covariate in a regression model that related changes in HRQOL scores to changes in the anchor showed no statistically significant effect ($p > 0.05$) of time separation on changes in HRQOL scores for each of the scales analyzed. Further, addition of “study effect” to the regression model showed no statistically significant differences in change scores between the two trials, supporting the idea of combining the two trials.

The mean change scores for the selected QLQ C-30 domains and corresponding differences between adjacent categories are presented in Tables 3 and 4, anchored by PS and weight change, respectively.

As an illustration, in Table 3, the first difference in PF mean change of adjacent categories is obtained as 3.6−(−5.3)=8.9 PF units, and the second is calculated as −5.3−(−9.7)=4.4 PF units, providing MID estimates for improvement and deterioration, respectively, similarly for weight change. For the PS results, the 95% CI for the difference in mean of change scores did not include zero, suggesting

Table 2 Summary statistics for HRQOL scales at baseline and at T1 and T2, cross-sectional correlation estimates for HRQOL scores with PS, and correlations between HRQOL change scores with changes in both anchors

	HRQOL summary statistics at baseline			Follow-up HRQOL, mean (SD)		Cross-sectional correlation (HRQOL and PS)		Correlation between changes in HRQOL and changes in anchor	
	Mean (SD)	Median	IQR	T ₁	T ₂	T ₁	T ₂	PS	Weight change
Physical	72.4 (23.7)	80.0	33.3	73.0 (24.2)	66.5 (25.1)	-0.37	-0.44	-0.17	0.19
Role	58.7 (33.2)	66.7	50.0	61.3 (34.5)	56.8 (33.0)	-0.39	-0.31	-0.20	0.15
Social	75.8 (28.6)	83.3	33.3	74.9 (28.1)	72.7 (28.1)	-0.23	-0.31	-0.11	0.15
GHS	58.7 (22.1)	58.3	25.0	59.4 (21.7)	58.8 (21.1)	-0.29	-0.30	-0.14	0.10
Fatigue	37.2 (26.4)	33.3	33.4	37.7 (26.6)	43.9 (26.8)	0.34	0.33	0.21	-0.20
Pain	31.6 (31.3)	16.7	50.0	31.6 (31.0)	29.9 (30.6)	0.30	0.24	0.13	0.03

IQR interquartile range

statistically significant differences between the “improvement” and “no change” groups for all scales except for SF. For “no change”–“deterioration” comparisons, only SF showed a statistically significant difference. No statistically significant difference was observed between the “weight gain” and “no change” groups for all scales, while for “no change” versus “weight loss”, all scales except RF and GHS showed statistically significant differences.

Table 5 displays the anchor-based MID estimates adjacent to the distribution-based MID estimates. Since the SEM, 0.5 SD and 0.2 SD estimates at T₁ and T₂ were very similar, and not systematically different across the different scales and across the anchors, only results at T₁ based on PS were reported.

Discussion

The aim of our study was to determine the magnitude of difference in scores in selected EORTC QLQ-C30 scales

that represents the MID in palliatively treated NSCLC patients. Our approach was to link changes in HRQOL scores to groups known to have changed in terms of clinically relevant anchors, in this case, PS and weight change. In general, the mean changes in HRQOL within each anchor-defined group were in the expected direction.

It is notable that while not being definitive, the MID estimates differed somewhat in size across scales and when anchoring with PS, MIDs for improvement tended to be larger than MIDs for deterioration. The former tended to be closer to the SEM, while the latter tended to be closer to the 0.2 SD estimates. This provides further evidence that the 0.5 SD may represent a “medium” effect size [7], whereas 1 SEM may approximate a threshold for defining the MID [8]. In line with our results, Samsa et al. [9] suggest that 0.2 SD may provide a better estimate of MID than 0.5 SD.

The suggestion that a larger degree of change may be required to be meaningful when a patient is improving compared to worsening contrasts with a number of studies that have reported higher MID estimates for deterioration

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

Scale	Improvement of 1 category, n [43–57]	No change, n [295–354]	Deterioration of 1 category, n [72–108]	Difference in mean change (95% CI)	
				Improvement	Deterioration
Physical	3.6 (20.7)	-5.3 (17.4)	-9.7 (23.3)	8.9 (3.2, 14.7) ^a	4.4 (-0.5, 9.2)
Role	9.7 (33.0)	-4.3 (26.4)	-8.9 (32.9)	14.0 (5.3, 22.8) ^a	4.6 (-2.5, 11.7)
Social	3.3 (22.8)	-1.2 (22.8)	-8.5 (29.0)	4.5 (-2.0, 10.9)	7.3 (2.0, 12.6) ^a
GHS	8.2 (21.6)	-0.9 (19.7)	-4.5 (24.6)	9.1 (3.4, 14.7) ^a	3.6 (-0.9, 8.2)
Fatigue	-7.5 (27.1)	6.6 (24.3)	12.3 (31.7)	-14.1 (-21.1, -7.2) ^a	-5.7 (-11.3, 0.0)
Pain	-15.8 (34.0)	0.0 (26.7)	-3.1 (27.2)	-15.8 (-23.6, -7.9) ^a	3.1 (-5.0, 5.6)

Physical and role 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. 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)

^a Differences that are statistically significant

Table 4 Weight change—mean (SD) of HRQOL change scores in the three anchor-defined groups and the difference in mean change scores (95% CI) between adjacent groups

Scale	5-<20% gain, <i>n</i> [38–44]	No change, <i>n</i> [295–367]	5-<20% loss, <i>n</i> [80–109]	Difference in mean change (95%) CI	
				Improvement	Deterioration
Physical	1.1 (18.4)	-4.3 (17.3)	-10.5 (23.5)	5.4 (-0.4, 11.2)	6.2 (1.5, 10.8) ^a
Role	3.5 (28.3)	-3.6 (25.5)	-8.9 (36.2)	7.1 (-1.6, 15.9)	5.3 (-1.6, 12.8)
Social	5.8 (21.5)	-0.9 (21.9)	-9.7 (31.7)	6.7 (-0.3, 13.6)	8.8 (3.6, 14.2) ^a
GHS	3.9 (20.6)	-0.2 (20.0)	-3.9 (25.8)	4.1 (-2.3, 10.5)	3.7 (-1.0, 8.4)
Fatigue	-1.1 (26.9)	3.9 (24.9)	14.7 (30.7)	-5.0 (-13.0, 3.1)	-10.8 (-16.5, -5.2) ^a
Pain	-2.8 (29.1)	-0.8 (25.6)	-8.0 (33.1)	2.0 (-10.4, 6.4)	7.2 (1.2, 13.0) ^a

Physical and role 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. Difference in mean change refers to the difference in mean of HRQOL change scores between the “weight gain” and “no change” (improvement) and between the “no change” and “weight loss” (deterioration)

^aDifferences that are statistically significant

compared to improvement [14, 15, 26]. One possible explanation of our findings is that physicians may have misclassified the PS of patients, particularly those who they thought had stable PS. Our results suggest that patients classified as having not changed in PS had actually deteriorated in HRQOL scores. However, there is no valid a priori reason to suggest that there are differences in the way physicians assessed PS in our study relative to other studies. If a subconscious bias exists that makes it more likely for physicians to report PS as stable or worsening, rather than improving, then a larger MID for improvement would be found by our anchoring method. This is consistent with optimism bias, the same cognitive bias which may lead to the opposite result (a larger MID for deterioration) when patient ratings are used to anchor MID [27]. Our results are supported by the relatively large sample size available in this study, but further investigations in other cancer sites are required to confirm our results.

It is also possible that the differences in MIDs for improvement and deterioration were due merely to sampling variation. Our findings are based on the largest

samples in the literature to date for determining MIDs from HRQOL change scores and particularly for considering improvement versus deterioration. Nevertheless, when the 95% CIs are taken into account, there is considerable overlap in the MIDs for all scales for both improvement and deterioration. Thus, further studies of large samples of patients with cancers in other sites are warranted.

Due to the relatively weak correlations observed, we acknowledge that our anchors did not appear to work well for some of the subscales, e.g., SF and PA. It may be argued that such anchors may not be used for such scales. Other studies using other anchors [14, 15] have also found only moderately strong correlations of the anchors with the HRQOL scores; the reason(s) is (are) unknown. For interpretation, it could be recommended to augment our anchor-based MID estimates with results from one of the distribution-based approaches by considering only those anchor-based MID estimates (see Table 5) at least equal to 0.2 SD [13], which is a “small effect” [7].

The clinical significance of weight gain/loss is not well established. While weight gain in some patients is a

Table 5 Anchor-based MID estimates compared with distribution-based MID estimates

	Anchor-based MID estimates		Distribution-based MID estimates		
	Improvement with PS, weight gain	Deterioration with PS, weight loss	SEM	0.5SD	0.2SD
Physical	9, 5	4, 6	7	12	5
Role	14, 7	5, 5	14	17	6
Social	5, 7	7, 9	10	14	6
GHS	9, 4	4, 4	9	11	4
Fatigue	14, 5	6, 11	11	13	5
Pain	16, 2	3, 7	12	16	6

Anchor-based MID estimates are based on differences between T₁ and T₂. Distribution-based estimates are based on HRQOL scores at T₁ using PS as an anchor

positive sign of improving physical condition, in some patients weight gain can be due to a buildup of ascites, which may in fact reflect increasing cancer activity. Similarly, weight loss may paradoxically reflect health improvement, for instance, discrete ascites or oedema that improved with treatment. This complicates the relationship between weight change and the changes in HRQOL scores. Therefore, results for scales (e.g., pain) exhibiting weak correlations with this anchor should be interpreted with caution.

The correlations between changes in either anchor (PS or weight) and HRQOL were not strong. The functional relationship between these changes is unknown and can be complex. Further, correlation based on changes in scores and anchor is likely to be smaller than cross-sectional correlations due to the measurement error in both anchor and HRQOL measures at both time points.

The changes that we are calling MIDs are based on the definitions and clinical anchors that we have applied, i.e., they are not “absolute” but are “relative” to the clinical anchors we used, as is always the case for MIDs based on the anchor-based approach. Indeed, issues about whether changes in HRQOL scores corresponding to changes in these anchors represent the MID, rather than just an important, anchor-dependent difference can be raised. This is an important issue which requires further research.

There is limited information about whether the value of the MID is stable across the continuum of illness, for example, can a change in PS from 0 to 1 be considered the same as a change from 3 to 4, when correlating with changes in HRQOL scores? Patients in our trials had to have a fairly good PS to get into the trial (see Table 1); therefore, we did not have the data to address such an important issue.

The patients in our data showed a predominance of good functioning and low symptoms at baseline. Since MIDs can vary across the spectrum of the EORTC QLQ-C30 scores, it is conceivable that our results could be different if we had predominantly worse patients. Being retrospective in nature, our analysis was restricted to using only WHO PS and weight loss, which were deemed credible anchors for the selected scales. We could possibly have used different anchors if available or could still have used the anchors we considered, together with any other credible anchors if available. However, using different anchors or anchor types (e.g., subjective vs. objective or prospective vs. retrospective) can also lead to different conclusions regarding estimates of the MID [8].

In conclusion, our findings provide estimates of MIDs for NSCLC patients in a selected subset of the EORTC QLQ-C30 scales. Differences in MIDs for improvement and deterioration were observed albeit not definitive. Our MID estimates are in line with the findings of Osoba et al.

[14]. These estimates generally agree with the estimates of 5 to 10 units of the QLQ-C30 scales we tested and as proposed by Osoba et al. [14] and King [16]. We suggest that they may be used as guidance for clinicians and researchers to classify patients as improved or deteriorated in HRQOL and symptoms over time, and thence to determine the proportion of patients benefiting from treatment. They can also be used for sample size determination in the design of future clinical trials.

Acknowledgements This study was funded by an unrestricted academic grant from the Pfizer Foundation. We thank the EORTC clinical Lung Group and their clinical investigators and all the patients who participated in these trials.

Conflict of interest statement The authors indicated no potential conflicts of interest.

References

1. Jaeschke R, Singer J, Guyatt GH (1989) Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 10:407–415
2. Guyatt GH, Osoba D, Wu AW, Clinical Significance Consensus Meeting Group et al (2002) Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 77:371–383
3. Schünemann HJ, Guyatt GH (2005) Goodbye M(C)ID! Hello MID, where do you come from? *Health Serv Res* 40:593–597
4. Schünemann HJ, Puhan M, Goldstein R, Jaeschke R, Guyatt GH (2005) Measurement properties and interpretability of the Chronic Respiratory Disease Questionnaire (CRQ). *COPD* 2:81–89
5. King MT, Stockler MS, Cella D, Osoba D, Eton D, Thompson J, Eisenstein A (2010) Meta-analysis provides evidence-based effect sizes for a cancer-specific quality of life questionnaire, the FACT-G. *J Clin Epidemiol* 63:270–281
6. Lydick F, Epstein RS (1993) Interpretation of quality of life changes. *Qual Life Res* 2:221–226
7. Cohen J (1988) *Statistical power analysis for the behavioural sciences*. Academic, New York
8. Crosby RD, Kolotkin RL, Williams GR (2003) Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 56:395–407
9. Samsa G, Edelman D, Rothman ML et al (1999) Determining clinically important differences in health status measures: a general approach with illustration to the health utilities index mark II. *Pharmacoeconomics* 15:141–155
10. Norman GR, Sloan JA, Wyrwich KW (2003) Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 41:582–592
11. Beaton DE (2003) Simple as possible? Or too simple? Possible limits to the universality of the one-half standard deviation (comment). *Med Care* 41:593–596
12. Farivar SS, Kiu H, Hays RD (2004) 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* 4(5):521–529
13. Cella D, Eton DT, Lai J, Peterman AH, Merkel DE (2002) 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* 24:547–561

14. Osoba D, Rodrigues G, Myles J, Zee B, Pater J (1998) Interpreting the significance of changes in health related quality-of-life scores. *J Clin Oncol* 16:139–144
15. Ringash J, O'Sullivan B, Bezjak A, Redelmeier DA (2007) Interpreting clinically significant changes in patient-reported outcomes. *Cancer* 110(1):196–202
16. King MT (1996) The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. *Qual Life Res* 5:555–567
17. Cocks K, King MT, Velikova G, Fayers PM, Brown JM (2008) 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* 44:1793–1798
18. Smit EF, van Meerbeeck JPAM, Lianes P, Debruyne C, Legrand C, Schramel F et al (2003) Three-arm randomized study of two cisplatin-based regimens and paclitaxel plus gemcitabine in advanced non-small cell lung cancer: a phase III trial of the European Organization for Research and Treatment of Cancer Lung Cancer Group-EORTC 08975. *J Clin Oncol* 21:3909–3917
19. Giacconne G, Splinter TAW, Debruyne C, Khot GS, Lianes P, van Zandwijk N et al (1998) Randomized study of paclitaxel–cisplatin versus cisplatin–teniposide in patients with advanced non-small cell lung cancer. *J Clin Oncol* 16:2133–2141
20. Fayers P, Aaronson N, Bjordal K, Groenvold M, Curran D, Bottomley A (2001) EORTC QLQ-C30 Scoring Manual, 3rd edn. EORTC Quality of life Study Group, Brussels
21. Ringash J, Bezjak A, O'Sullivan B, Redelmeier D (2004) Interpreting small differences in quality of life: the FACT-H&N in laryngeal cancer patients. *Qual Life Res* 13(4):721–729
22. Cella D, Eton DT, Fairclough DL, Bonomi P, Heyes AE, Silberman C et al (2002) What is a clinically meaningful change on the Functional Assessment of Cancer Therapy-Lung (FACT-L) Questionnaire? Results from Eastern Cooperative Oncology Group (ECOG) Study 5592. *J Clin Epidemiol* 55:285–295
23. National Cancer Institute (2003) Cancer Therapy Evaluation Program, Common Terminology Criteria for Adverse Events (CTCAE), version 3.0, DCTD, NCI, NIH, DHHS 2003. <http://ctep.cancer.gov>. Accessed 28 Sept 2010
24. Revicki D, Hays RD, Cella D, Sloan J (2008) Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 61:102–109
25. Hjermstad MJ, Fossa SD, Bjordal K, Kaasa S (1995) Test/retest study of the European organization for research and treatment of cancer core quality of life questionnaire. *J Clin Oncol* 13:1249–1254
26. Cella D, Hahn EA, Dineen K (2002) Meaningful change in cancer-specific quality of life scores: differences between improvement and worsening. *Qual Life Res* 11(3):207–221
27. Weinstein ND (1989) Optimistic biases about personal risks. *Science* 246:1232–1233