European Journal of Cancer 188 (2023) 171–182

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Original Research

Minimally important differences for interpreting EORTC QLQ-C30 change scores over time: A synthesis across 21 clinical trials involving nine different cancer types

Jammbe Z. Musoro [a,*], Corneel Coens [a], Mirjam A.G. Sprangers [b,c], Yvonne Brandberg [d], Mogens Groenvold [e], Hans-Henning Flechtner [f], Kim Cocks [g], Galina Velikova [h,v], Linda Dirven [i,j], Elfriede Greimel [k], Susanne Singer [l,m], Katarzyna Pogoda [n], Eva M. Gamper [o], Samantha C. Sodergren [p], Alexander Eggermont [q,r], Michael Koller [s], Jaap C. Reijneveld [t], Martin J.B. Taphoorn [i,j], Madeleine T. King [u], Andrew Bottomley [a], on behalf of the EORTC Melanoma, Breast, Head and Neck, Genito-urinary, Gynecological, Gastro-intestinal, Brain, Lung and Quality of Life Groups

[a] *European Organisation for Research and Treatment of Cancer (EORTC), Brussels, Belgium*
[b] *Amsterdam UMC Location University of Amsterdam, Medical Psychology, Amsterdam, The Netherlands*
[c] *Cancer Center Amsterdam, Cancer Treatment and Quality of Life, Amsterdam, The Netherlands*
[d] *Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden*
[e] *Department of Public Health, University of Copenhagen, and Bispebjerg Hospital, Copenhagen, Denmark*
[f] *Clinic for Child and Adolescent Psychiatry and Psychotherapy, University of Magdeburg, Magdeburg, Germany*
[g] *Adelphi Values, Bollington, Cheshire, UK*
[h] *Leeds Institute of Medical Research at St James's, University of Leeds, St James's University Hospital, Leeds, UK*
[i] *Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands*
[j] *Department of Neurology, Haaglanden Medical Center, The Hague, The Netherlands*
[k] *Medical University Graz, Graz, Austria*
[l] *Institute of Medical Biostatistics, Epidemiology and Informatics, Division of Epidemiology and Health Services Research, University Medical Centre Mainz, Germany*
[m] *University Cancer Centre Mainz, Germany*
[n] *Departmenf of Breast Cancer and Reconstructive Surgery, Maria Sklodowska-Curie National Research Institute of Oncology, Warsaw, Poland*
[o] *Innsbruck Institute of Patient-centered Outcome Research (IIPCOR), Innsbruck, Austria*
[p] *School of Health Sciences, University of Southampton, Southampton, UK*
[q] *Princess Máxima Center, Utrecht and University Medical Center Utrecht, The Netherlands*
[r] *Comprehensive Cancer Center Munich, Technical University Munich & Ludwig Maximiliaan University, Munich, Germany*
[s] *Center for Clinical Studies, University Hospital Regensburg, Regensburg, Germany*

[*] *Corresponding author*: Quality of Life Department, European Organization for Research and Treatment of Cancer, 83/11 Avenue E. Mounier, 1200 Brussels, Belgium.
E-mail address: jammbe.musoro@eortc.org (J.Z. Musoro).

https://doi.org/10.1016/j.ejca.2023.04.027

[t] Amsterdam University Medical Centers, location VU University Medical Center, Department of Neurology Brain Tumor Center, Amsterdam, The Netherlands
[u] University of Sydney, Faculty of Science, School of Psychology, Sydney, NSW, Australia
[v] Leeds Cancer Centre, Leeds Teaching Hospitals NHS Trust, St James's University Hospital, Leeds, UK

Received 15 February 2023; Received in revised form 27 April 2023; Accepted 27 April 2023
Available online 7 May 2023

KEYWORDS
Patient-reported outcomes (PRO);
Health-related quality of life (HRQoL);
EORTC QLQ-C30;
Minimally important difference (MID);
Group-level change;
Cancer

Abstract ***Introduction:*** Early guidelines for minimally important differences (MIDs) for the EORTC QLQ-C30 proposed ≥10 points change as clinically meaningful for all scales. Increasing evidence that MIDs can vary by scale, direction of change, cancer type and estimation method has raised doubt about a single global standard. This paper identifies MID patterns for interpreting group-level change in EORTC QLQ-C30 scores across nine cancer types.
Methods: Data were obtained from 21 published EORTC Phase III trials that enroled 13,015 patients across nine cancer types (brain, colorectal, advanced breast, head/neck, lung, mesothelioma, melanoma, ovarian, and prostate). Anchor-based MIDs for within-group change and between-group differences in change over time were obtained via mean change method and linear regression, respectively. Separate MIDs were estimated for improvements and deteriorations. Distribution-based estimates were derived and compared with anchor-based MIDs.
Results: Anchor-based MIDs mostly ranged from 5 to 10 points. Differences in MIDs for improvement vs deterioration, for both within-group and between-group, were mostly within a 2-points range. Larger differences between within-group and between-group MIDs were observed for several scales in ovarian, lung and head/neck cancer. Most anchor-based MIDs ranged between 0.3 SD and 0.5 SD distribution-based estimates.
Conclusions: Our results reinforce recent claims that no single MID can be applied to all EORTC QLQ-C30 scales and disease settings. MIDs varied by scale, improvement/deterioration, within/between comparisons and by cancer type. Researchers applying commonly used rules of thumb must be aware of the risk of dismissing changes that are clinically meaningful or underpowering analyses when smaller MIDs apply.

1. Introduction

The past years have witnessed a growth in the use of patient-reported outcomes (PROs) in cancer clinical trials to support informed claims about treatment risks, benefits, safety, and tolerability [1,2]. This growth coincides with increasing efforts to enforce various standards to improve, among others, the collection, reporting, analysis and interpretation of PRO data in cancer clinical trials [3–5]. When interpreting PRO results, it is crucial to understand the degree of change in PRO scores that a patient perceives as clinically relevant.

The notion of minimally important difference (MID) is one of several frameworks that help attach clinically meaningful interpretations to PRO data. MID is defined as 'smallest difference in score in the domain of interest that patients perceive as important, either beneficial or harmful, and which would lead the clinician to consider a change in the patient's management' [6]. This paper examines MID guidelines for interpreting PRO results based on the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) when comparing groups of patients in cancer clinical trials.

The EORTC QLQ-C30 is one of the most widely used PRO measures for assessing patients' health-related quality of life (HRQoL) in cancer research [7]. Early MID guidelines suggested differences of ≥10 points as clinically relevant for all EORTC QLC-C30 scales [8,9]. However, increasing evidence that clinically meaningful thresholds can differ by scale, direction of change, anchor, cancer type, as well as estimation method, has raised doubt about the generalisability of this single global standard [10,11]. A way forward is to adopt guidelines that advocate a more nuanced, yet practical, strategy to clinical relevance beyond a single threshold [10]. In this light, the EORTC Quality of Life Group funded the MID project to gather empirical evidence on MID patterns across EORTC QLQ-C30 scales and across different cancer types, using data from published cancer clinical trials [12].

The main goal of this paper is to present an overview of estimated MID values for interpreting group-level change of EORTC QLQ-C30 scores over time that were derived across nine cancer types (brain, colorectal, advanced breast, head/neck, lung, mesothelioma, melanoma, ovarian, and prostate) [13–15, 17–20]. Specifically, MID patterns will be identified by scales, direction of change and intended application (for within vs between group comparison) per cancer type. A secondary goal is to compare our MID estimates to previously published MID guidelines for EORTC QLQ-C30 change scores.

2. Materials and methods

2.1. Data

The study data were obtained from 21 published EORTC phase III trials that enroled 13,015 patients in total, across nine different cancer types [13–20]. This included three brain cancer trials (total pooled sample size, n = 1697), 3 colorectal cancer trials (n = 1491), 2 advanced breast cancer trials (n = 723), 2 head/neck cancer trials (n = 808), 1 lung cancer trial (n = 480), 1 malignant pleural mesothelioma trial (n = 250), 3 melanoma trials (n = 3595), 4 ovarian cancer trials (n = 2034), and 2 prostate cancer trials (n = 1937). All trials collected HRQoL data as measured with the EORTC QLQ-C30 at baseline, and multiple time points during and after treatment. The data were pooled and analysed by cancer type, except for the lung cancer and mesothelioma trials that were combined and analysed as one because of their common respiratory problems, impact on patients' activities and other aspects of HRQoL [20].

2.2. EORTC QLQ-C30

This HRQoL questionnaire is designed for any cancer population and comprises 30 distinct questions that are scored into 15 scales [21]. These include five functional scales (physical, role, emotional, social, and cognitive), eight symptom scales (pain, fatigue, nausea and vomiting, insomnia, appetite loss, constipation, diarrhoea, dyspnoea), a financial difficulties scale and a global health status/quality of life scale. The reliability and validity of this questionnaire is highly consistent across different language and cultural groups [21]. For consistency in interpretation, we deviated from the standard scoring procedure by scoring all scales such that 0 represents the worst possible score and 100 the best possible score. The financial difficulties scale was omitted from the study due to lack of suitable anchors.

2.3. Data analysis

MIDs were derived using two main approaches. The primary approach used anchor-based methods where thresholds for meaningful change were established by linking QLQ-C30 scale scores to independent outcomes with known clinical relevance. The secondary approach was the distribution-based approach which relies on the dispersion of QLQ-C30 scale scores [22]. Anchor-based MIDs were estimated using change scores of both the anchors and the QLQ-C30 scales computed across all pairwise assessment time points, and then combined into one dataset. That is, if a subject was measured at three time points t_1, t_2 and t_3, change scores were computed between t_1 & t_2, t_1 & t_3 and t_2 & t_3. Change scores were only calculated if both QLQ-C30 and anchor data were available at a given pair of time points. Distribution-based estimates were calculated using baseline data, that is, data at the time point before or on the first day of treatment administration. Distribution-based estimates were also evaluated and compared to the anchor-based MIDs. Details on these methods have been described in our previous publications [13–20]. Below we provide a summary.

2.3.1. Anchor-based approach

i. *Clinical anchor selection and definition of anchor change groups*
Clinical anchors, such as performance status (PS) and common terminology criteria for adverse events (CTCAE), were screened based on correlation strength with a particular EORTC QLQ-C30 scale. PS was scored between 0 (no symptoms of cancer) and 4 (bedbound), while the CTCAEs were graded between 0 (no toxicity) and 4 (life-threatening). Priority was given to anchors with correlations ≥|0.30| and where attainable, anchors with higher correlations were targeted [22]. The computed correlations have been published previously by cancer type [13–20]. Identified anchors were further evaluated for clinical plausibility (i.e. if there is a clinical basis for the relationship between anchor and scale) by international clinical and HRQoL experts to ensure interpretable results. When available, multiple anchors were used per scale. Three anchor-change groups were formed: deteriorated by one anchor category, improved by one anchor category and no change over time. Patients with change scores ≥2 anchor categories were not used for MID estimation since they were considered to be clearly above the 'minimal' expected change.

ii. *MIDs for within-group change and between-group differences in changes over time*

- *MIDs for within-group change over time:* Within-group change was defined as the change within the same groups of patients assessed at two time points. The associated MID was estimated by the mean change in QLQ-C30 scores of patients who improved or deteriorated on the clinical anchor, respectively (mean change method) [12]. For each treatment arm in a trial, a mean HRQOL change score over time that is ≥ the *within-group* MID would be considered clinically meaningful.
- *MIDS for between-group differences in change over time:* Between-group difference in change over time was defined

as the difference between two groups in the change within group assessed at two time points. The associated MID was estimated using linear regression models with the QLQ-C30 change score as outcome and a binary anchor indicator of 'stable' vs 'deterioration when modelling deterioration (excluding observations indicating improvement) and vice versa. The MIDs for deterioration and improvement correspond to estimated slopes of the 'deterioration' and 'improvement' anchor covariate, respectively. In a trial, a difference between the mean HRQOL change score in an experimental treatment group compared to a control group that is ≥ the *between-group* MID would be considered clinically meaningful.

- We conducted sensitivity analysis to assess whether MIDs varied by potential confounding factors as age, gender, and trial. We included each factor separately and their interaction with the binary anchor indicator in a regression model.

When multiple MIDs (in case of multiple anchors) were estimated for the same EORTC QLQ-C30 scale, they were summarised into a single value by calculating a correlation-weighted average. Weights were constructed so that anchors having stronger correlations with a given scale contributed more to the single MID estimate [23].

2.3.2. Distribution-based methods

For each cancer type, three proportions of a standard deviation (0.2 SD, 0.3 SD, 0.5 SD) were calculated using only baseline data. Additionally, the standard error of measurement (SEM) was also calculated as $SD\sqrt{(1-r)}$, using SD at baseline. The test-retest reliability estimates (r) for the QLQ-C30 scales were obtained from Hjermstad et al. [24]. Although these distribution-based estimates have previously been considered relevant to determining MIDs [22,25], there is currently no consensus on which estimate best approximates the MID. Since 0.2 SD and 0.5 SD reflect a small and medium effect size, respectively [26], differences < 0.2 SD are likely to be below the MID while differences significantly above 0.5 SD are likely to be above the MID [27].

In addition, within-group effect sizes (ES) were computed within each anchor-change group by taking the mean of the QLQ-C30 change scores divided by the SD of the change scores. Based on Cohen's guidelines [26], only mean changes with ES ≥0.2 and < 0.8 were considered appropriate for inclusion as anchor-based MID estimates since ESs < 0.2 reflect clinically unimportant changes, while ESs ≥0.8 are beyond 'minimally' important.

3. Results

3.1. Patient characteristics

A summary of patients' demographic/clinical characteristics and the distribution of EORTC QLQ-C30 scale scores at baseline are presented by cancer type in Supplementary Tables 1 and 2, respectively. A more detailed description of patient baseline characteristics has previously been published by disease type [13–20].

3.2. Anchor-based MID estimates

3.2.1. Clinical anchor selection

Selection of clinical anchors depended on available data within the various disease-specific trial databases. Hence, the final list of anchors that was retained varied by QLQ-C30 scale and by disease type. For instance, in the melanoma studies, at least two anchors were found for each of the 14 scales, while in the prostate cancer studies at least one clinical anchor was identified for only seven of the 14 scales. As shown in Fig. 1, the retained anchors across the various disease types were mainly PS and CTCAEs such as fatigue, nausea/vomiting and gastrointestinal symptoms.

3.2.2. Patterns of anchor-based MID estimates

MIDs for within-group change along with their 95% confidence intervals (obtained via the mean change method) are plotted in Fig. 1. Generally, MIDs varied by QLQ-C30 scale, anchor, direction of change (i.e. improvement vs deterioration) and by cancer type. Where available, multiple anchors per scale provided greater confidence in the appropriateness of the MID estimates, which were often close to each other. Relatively wider CIs for MID estimates were mainly observed for cancer types and/or anchor/QLQ-C30 pairs with smaller sample sizes. Results for prostate and head/neck cancer are omitted in Fig. 1 because only one anchor was available for most scales. Similar patterns were observed for the between-group MIDs from the linear regression approach. No indications of deviation from linearity were noticed during for linear regression models (results not shown).

Summary of anchor-based MIDs across different cancer types for within-group and between-group differences in change over time are presented in Tables 1 and 2, respectively. Both single values and range of MIDs (for scales with multiple anchors) are presented. A weighted MID average is also presented for those scales where a range of MIDs was available. Fig. 2 plots scale-specific single value MIDs (absolute values) for within and between-group change and separately for improvement vs deterioration, per cancer type.

In general, most MIDs were within a 5–10 points range. The smallest MID of 3 points was observed for between-group difference in social function change scores for patients with prostate cancer. Moreover, differences in MIDs for improvement vs deterioration, based on both within-group and between-group analyses, were within a 2-points range for most scales and for most cancer types. However, MIDs for improvements tended to be larger than those for deterioration

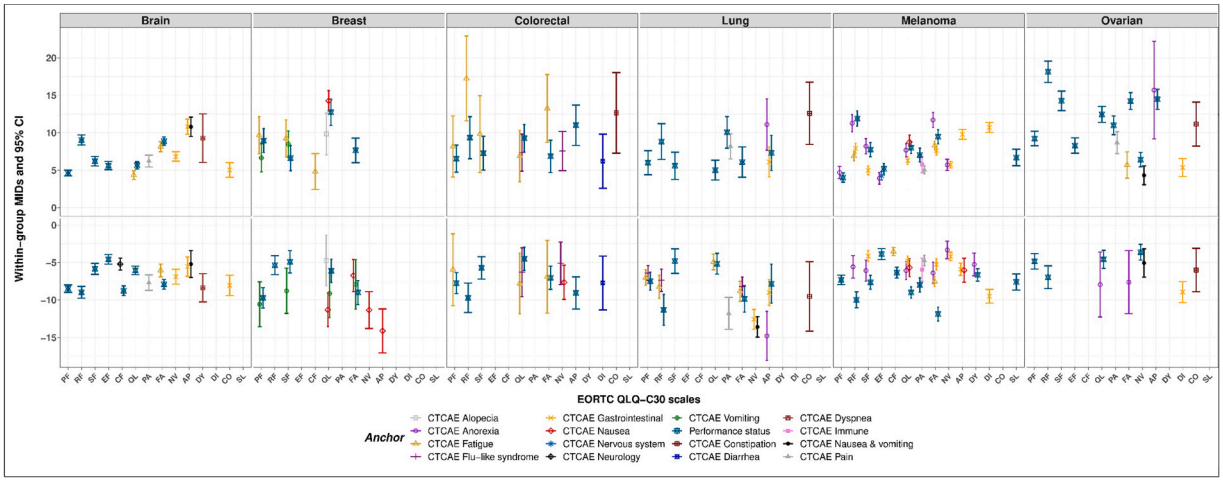

Fig. 1. Within-group MIDs (95% CI) for improvement (upper half) and deterioration (lower half) in EORTC QLQ-C30 scales across multiple anchors by cancer type. MIDs were obtained from the mean change method and are available only for scales with at least 1 suitable anchor or with effect size ≥0.2 and < 0.8 within the 'deteriorate' and 'improve' groups, respectively. Lung cancer and mesothelioma trials were pulled and analysed together. Results for prostate and head/neck cancer are omitted because only one anchor was available for most scales.
Abbreviations: AP, appetite loss; CF, cognitive functioning; CO, constipation; DI, diarrhoea; DY, dyspnoea; EF, emotional functioning; FA, fatigue; NV, nausea/vomiting; PA, pain; PF, Physical functioning, QL, global health/quality of life; RF, role functioning; SF, social functioning; SL, sleep disturbance; CTCAE, common terminology criteria for adverse events; CI confidence interval. Deteriorate = worsened by 1 anchor category, no change = no change in anchor category and improve = improved by 1 category.

for most scales in ovarian and colorectal cancer, for both within-group and between-group analyses and only for within-group change for melanoma patients. Further, in lung and head/neck cancer, MIDs for deteriorating scores for most scales were larger based on within-group change compared to between-group difference in change scores. The largest MID differences between improvement vs deterioration were observed mostly for within-group change in ovarian cancer for 5 scales: physical functioning (9 vs −5), role functioning (18 vs −7), global health status (13 vs −6), fatigue (12 vs −5) and constipation (11 vs −6). There were no systematic differences in MID values between the functioning vs symptom scales.

3.3. Distribution-based MID estimates

Distribution-based MID estimates for selected QLQ-C30 scales have been previously published by cancer type [13–20]. Estimates for all 14 scales considered in this study are collectively presented in Supplementary Fig. 1 (lower half) and Supplementary Tables 3a and 3b. In general, distribution-based MIDs ranged between 4 and 11 points across all scales per disease type. The median (range) was for *0.2 SD:* 5 (1–7), *0.3 SD*: 7 (2–10), *0.5 SD*: 11 (3–17) and *1 SEM*: 9 (4–15).

3.4. Comparison of anchor-based and distribution-based MID estimates

Overall, most anchor-based MIDs for the QLQ-C30 scales across the different cancer types were ≥*0.2 SD*, and tended to range between *0.3 SD* and *0.5 SD*. In brain cancer, breast cancer and melanoma, most anchor-based MIDs were closest to *0.3 SD or 1 SEM*. In ovarian cancer, anchor-based MIDs for improvement for most scales were closer to *0.5 SD*, whereas those for deterioration tended to range from *0.2 SD* to *0.3 SD*. In prostate cancer, with the exception of the diarrhoea scale, anchor-based MIDs for improvement were closer to 0.3 SD, while those for deterioration mainly ranged between *0.3 SD* and *0.5 SD*.

3.5. Comparison with existing MID guidelines

Other existing MID guidelines for the QLQ-C30 include Osoba et al. [9] based on results from trials among patients with breast and small-cell lung cancer, Maringwa et al. [28,29] in lung and brain cancer, respectively, Kawahara et al. [30] in advanced breast cancer, and Cocks et al. [10,11] and King [8] in pooled data across multiple cancer sites. In general, our results are consistent with the guidelines provided by Osoba et al. [9] in that most MIDs were in a 5–10 points range [9]. Furthermore, our estimates were also in line with the more recent guidelines [10,11,28–30], showing that MID magnitudes differ for improvement vs deterioration (although these differences are relatively small for most scales) and also depend on the particular QLQ-C30 scale.

More recently, Kawahara et al. [30] published MIDs for interpreting the QLQ-C30 scores in Japanese patients with advanced breast cancer. They found similar between-group MIDs for the global quality of life, physical function, role function, social function, fatigue,

Table 1
Summary of anchor-based MIDs across different cancer types for within-group change over time.

	Brain		Breast		Colorectal		Head/neck		Lung		Melanoma		Ovarian		Prostate	
Scale	**Imp (+)**	**Det (-)**	**Imp (+)**	**Det (-)**	**Imp (+)**	**Det (-)**	**Imp (+)**	**Det (-)**	**Imp (+)**	**Det (-)**	**Imp (+)**	**Det (-)**	**Imp (+)**	**Det (-)**	**Imp (+)**	**Det (-)**
	n_1 = 24–406	n_1 = 23–468	n_1 = 50–179	n_1 = 54–250	n_1 = 37–178	n_1 = 56–240	n_1 = 37–181	n_1 = 28–208	n_1 = 65–267	n_1 = 70–354	n_1 = 211–910	n_1 = 213–768	n_1 = 26–434	n_1 = 39–424	n_1 = 90–363	n_1 = 60–452
	n_2 = 115–6457	n_2 = 102–6423	n_2 = 299–1383	n_2 = 189–1182	n_2 = 122–619	n_2 = 96–1142	n_2 = 83–538	n_2 = 34–596	n_2 = 241–907	n_2 = 185–1349	n_2 = 2248–7548	n_2 = 1041–4881	n_2 = 106–2098	n_2 = 129–1598	n_2 = 404–1556	n_2 = 249–2237
Physical functioning	5	-9	7 to 10 (9)	-11 to -10 (-10)	7 to 9 (8)	-8 to -6 (-7)	NA	-11	6	-8 to -6 (-7)	4 to 5 (4.3)	-7	9	-5	NA	-11
Role functioning	9	-9	NA	-6	10 to 18 (14)	-11	NA	-15	9	-11 to -7 (-9)	7 to 12 (10)	-10 to -6 (-8)	19	-7	4	-13
Social functioning	6	-6	7 to 9 (8)	-9 to -5 (-7)	8 to 10 (9)	-6	7	-8	6	-5	5 to 8 (7)	-8 to -4 (-6)	15	NA	4	-5
Emotional functioning	6	-5	NA	NA	NA	NA	NA	NA	NA	NA	4 to 5 (4.5)	-4	9	NA	NA	NA
Cognitive functioning	NA	-9 to -5 (-7)	5	-4	NA	NA	NA	NA	NA	NA	NA	-6 to -4 (-5)	NA	NA	NA	NA
Global health status	4 to 6 (5)	-6	10 to 14 (12)	-11 to -5 (-8)	7 to 10 (8)	-8 to -5 (-6)	9	-5	5	-5	6 to 9 (8)	-9 to -5 (-6)	13	-8 to -5 (-6)	NA	-7
Fatigue	8 to 9 (8.7)	-8 to -6 (-7)	8	-9 to -7 (-8)	8 to 14 (11)	-8 to -7 (-7)	NA	-15	6	-10 to -8 (-9)	8 to 12 (9)	-12 to -5 (-8)	6 to 15 (12)	-8 to -5 (-6)	NA	-9
Pain	6	-8	NA	NA	NA	NA	NA	NA	8 to 10 (9)	-12	5 to 7 (6)	-8 to -5 (-6)	9 to 11 (10)	NA	NA	-6
Nausea and vomiting	7	-7	NA	-11	8	-8 to -5 (-7)	NA	-6	NA	-14 to -13 (-13)	6	-6 to -3 (-4)	4 to 7 (5)	-5 to -4 (-4)	NA	NA
Dyspnea	9	-8	NA	NA	NA	NA	6	-7	NA	NA	NA	NA	NA	NA	NA	NA
Appetite loss	11 to 12 (11)	-5 to -4 (-5)	NA	-14	12	-10	NA	NA	6 to 11 (8)	-15 to -8 (-11)	10	-15 to -6 (-9)	15 to 16 (15)	NA	NA	NA
Diarrhoea	NA	NA	NA	NA	6	-8	NA	NA	NA	NA	11	-10	5	-9	14	-9
Sleep disturbance	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	7	-8	NA	NA	NA	NA
Constipation	5	-14 to -9 (-10)	NA	NA	13	NA	NA	NA	13	-10	NA	NA	11	-6	NA	NA

The within-group MIDs are derived from the mean change method
The MIDs within parenthesis are averages weighted by the correlations between change score of scale/anchor pairs.
The symptom scores were reversed to follow the functioning scales' interpretation, that is, 0 represents the worst possible score and 100, the best possible score; 'NA 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. Note that the lung cancer and mesothelioma trials were pooled and analysed together.
n_1 and n_2 can vary by anchor and EORTC QLQ-C30 scale.
Details on the overview of patient inclusion for the anchor-based method as well as the distribution of n_1 and n_2 have been previously published by disease type [13–20].
Abbreviations: Imp (+) = Improvement; Det (-) = Deterioration; n_1 = number of patients with at least 2 completed forms (i.e. patients who completed at least the EORTC QLQ-C30 on two dates and who had anchor data for corresponding dates, because at least 2 forms are needed to compute change scores); n_2 = number of repeated EORTC QLQ-C30 scales and anchor change scores across all patients.

Table 2
Summary of anchor-based MIDs across different cancer types for between-group difference in change over time.

	Brain		Breast		Colorectal		Head/neck		Lung		Melanoma		Ovarian		Prostate	
Scale	**Imp (+)**	**Det (-)**	**Imp (+)**	**Det (-)**	**Imp (+)**	**Det (-)**	**Imp (+)**	**Det (-)**	**Imp (+)**	**Det (-)**	**Imp (+)**	**Det (-)**	**Imp (+)**	**Det (-)**	**Imp (+)**	**Det (-)**
	n_1 = 24–406	n_1 = 23–468	n_1 = 50–179	n_1 = 54–250	n_1 = 37–178	n_1 = 56–240	n_1 = 37–181	n_1 = 28–208	n_1 = 65–267	n_1 = 70–354	n_1 = 211–910	n_1 = 213–768	n_1 = 26–434	n_1 = 39–424	n_1 = 90–363	n_1 = 60–452
	n_2 = 115–6457	n_2 = 102–6423	n_2 = 299–1383	n_2 = 189–1182	n_2 = 122–619	n_2 = 96–1142	n_2 = 83–538	n_2 = 34–596	n_2 = 241–907	n_2 = 185–1349	n_2 = 2248–7548	n_2 = 1041–4881	n_2 = 106–2098	n_2 = 129–1598	n_2 = 404–1556	n_2 = 249–2237
Physical functioning	5	-7	7 to 9 (8)	-10 to -8 (-9)	6. to 10 (8)	-7 to -4 (-6)	NA	-7	5	-4	4 to 5 (5)	-6	6.	-6	NA	-7
Role functioning	8	-9	NA	-4	8 to 14 (11)	-10	NA	-12	7	-8 to -4 (-6)	5 to 11 (8)	-11 to -8 (-9)	13	-10	5	-10
Social functioning	5	-6	6 to 7 (7)	-11 to -5 (-8)	7 to 8 (7)	-6	5	-8	5	-4	5 to 8 (7)	-8 to -4 (-6)	10	NA	3	-4
Emotional functioning	4	-4	NA	NA	NA	NA	NA	NA	NA	NA	3 to 5 (4)	-5	6	NA	NA	NA
Cognitive functioning	NA	-6 to -5 (-6)	4	-4	NA	NA	NA	NA	NA	NA	NA	-5 to -2 (-4)	NA	NA	NA	NA
Global health status	3 to 5 (4)	-6	8 to 11 (10)	- 13 to -6 (-10)	6	-9. to -7 (-8)	5	-7	4	-4	6 to 9 (7)	-7 to -5 (-6)	9	-8 to -7 (-7)	NA	-6
Fatigue	8	-8 to -7 (-7)	8	-8 to -6 (-7)	5 to 12 (9)	-7	NA	12	6	-6 to -5 (-6)	7 to 11 (9)	-10 to -5 (-8)	3 to 10 (8)	-11 to -7 (-8)	NA	-7
Pain	7	-6	NA	NA	NA	NA	NA	NA	7 to 10 (9)	-9	4 to 7 (5)	-8 to -5 (-7)	6 to 8 (7)	NA	NA	-5
Nausea and vomiting	6	-7	NA	-14	7	-7 to -5 (-6)	NA	-5 to -4 (-4)	NA	-9 to -7 (-8)	5 to 6 (5)	-4	4 to 5 (4)	-6 to -4 (-5)	NA	NA
Dyspnea	7	-8	NA	NA	NA	NA	7	-4	NA	NA	NA	NA	NA	NA	NA	NA
Appetite loss	9	-8 to -7 (-8)	NA	-18	10	-7	NA	NA	6 to 15 (10)	-9 to -5 (-7)	9	-15 to -6 (-10)	9 to 13 (11)	NA	NA	NA
Diarrhoea	NA	NA	NA	NA	8	-6	NA	NA	NA	NA	10	-10	6	-8	13	-9
Constipation	5	-16 to -7 (-10)	NA	NA	15	NA	NA	NA	13	-9	NA	NA	7	-8	NA	NA
Sleep disturbance	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	7	-7	NA	NA	NA	NA

The between-group MIDs were derived from the linear regression
The MIDs within parenthesis are averages weighted by the correlations between change score of scale/anchor pairs.
The symptom scores were reversed to follow the functioning scales' interpretation, that is, 0 represents the worst possible score and 100, the best possible score; 'NA' 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. Note that the lung cancer and mesothelioma trials were pooled and analysed together.
n_1 and n_2 can vary by anchor and EORTC QLQ-C30 scale.
Details on the overview of patient inclusion for the anchor-based method as well as the distribution of n_1 and n_2 have been previously published by disease type [13–20].
Abbreviations: Imp (+) = Improvement; Det (-) = Deterioration; n_1 = number of patients with at least 2 completed forms (i.e. patients who completed at least the EORTC QLQ-C30 on two dates and who had anchor data for corresponding dates, because at least 2 forms are needed to compute change scores); n_2 = number of repeated EORTC QLQ-C30 scales and anchor change scores across all patients.

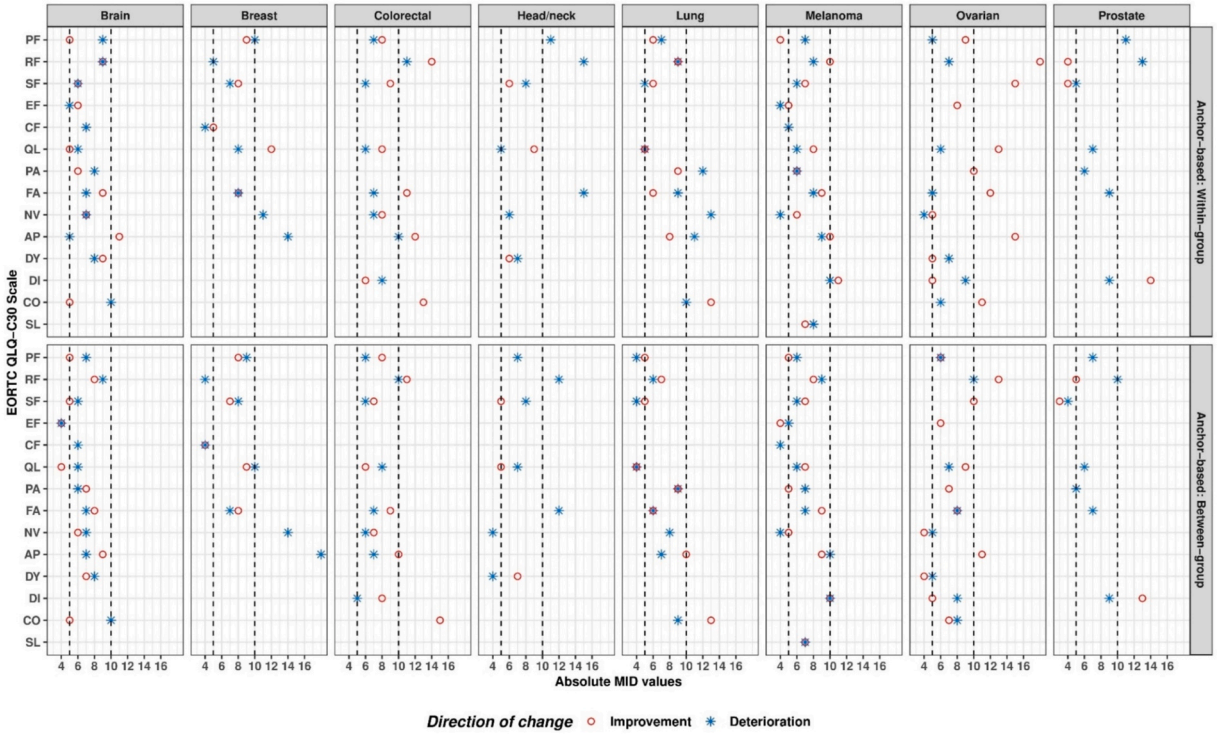

Fig. 2. Within-group (upper half) and between-group (lower half) MIDs for improvement and deterioration in EORTC QLQ-C30 scales by cancer types. The MIDs correspond to the single-value estimates presented in Tables 1 & 2. Estimates are available only for scales with at least 1 suitable anchor or with effect size ≥0.2 and < 0.8 within the 'deteriorate' and 'improve' anchor-change groups, respectively. Example on how to read plot: For brain cancer (see upper left part of figure), the within group MID for improvement on PF is 5 points and for deterioration in PF is 9 points in absolute values. Most MIDs were within a 5–10 points range (represented by the broken lines). Abbreviations: AP, appetite loss; CF, cognitive functioning; CO, constipation; DI, diarrhoea; DY, dyspnoea; EF, emotional functioning; FA, fatigue; NV, nausea/vomiting; PA, pain; PF, Physical functioning, QL, global quality of life; RF, role functioning; SF, social functioning; SL, sleep disturbance; CTCAE, common terminology criteria for adverse events. Deteriorate = worsened by 1 anchor category, no change = no change in anchor category and improve = improved by 1 category.

and appetite loss scales compared to our findings [18] that were based on data derived from mainly European patients with advanced breast cancer. Kawahara et al. [30] also reported MIDs for within-group deterioration that tended to be larger among the Japanese patients than among European patients [18], for example, role function (−17 versus −6) and social function (−12 versus −7).

Cocks et al. [10,11] have provided general guidance on MID selection for all 15 QLQ-C30 scales; for interpreting cross-sectional between-group differences [10] and within-group improvements vs deteriorations over time [11]. This was based on meta-analyses of published studies, pooling across 11 cancer types including breast, lung, head/neck, colorectal, prostate, haematological, gastrointestinal, brain, urology/kidney, testicular and gynaecological cancers. Supplementary Fig. 2 compares our single value MIDs per scale (taken from Tables 1 and 2 above) to the range of estimates identified by Cocks et al. for interpreting small within-group change [11] and small between-group difference [10]. Overall, our estimates were in the same range as those from Cocks et al. for most scales across the various disease sites.

4. Discussion

This study brings together MIDs for group-level interpretation of EORTC QLQ-C30 change scores over time across nine different cancer types (brain, colorectal, advanced breast, head/neck, lung, mesothelioma, melanoma, ovarian, and prostate) [13–20]. To date, this is the most comprehensive scrutiny of MID commonalities and differences for the QLQ-C30 across different cancer types by scale, direction of change (improvement vs deterioration) as well as for within vs between group comparisons. MIDs were derived mainly via anchor-based methods that targeted multiple anchors per EORTC QLQ-C30 scale to boost confidence in the plausibility of the MID estimates. Distribution-based estimates were also derived to support anchor-based estimates when available and for interpretation in the few cases where anchor-based MIDs were unavailable. The range of estimates from both distribution-based and anchor-based methods (Supplementary Fig. 1) generally supported the plausibility of our anchor-based MIDs.

Our results highlight the diversity in MID estimates because of the numerous possible anchors, the various

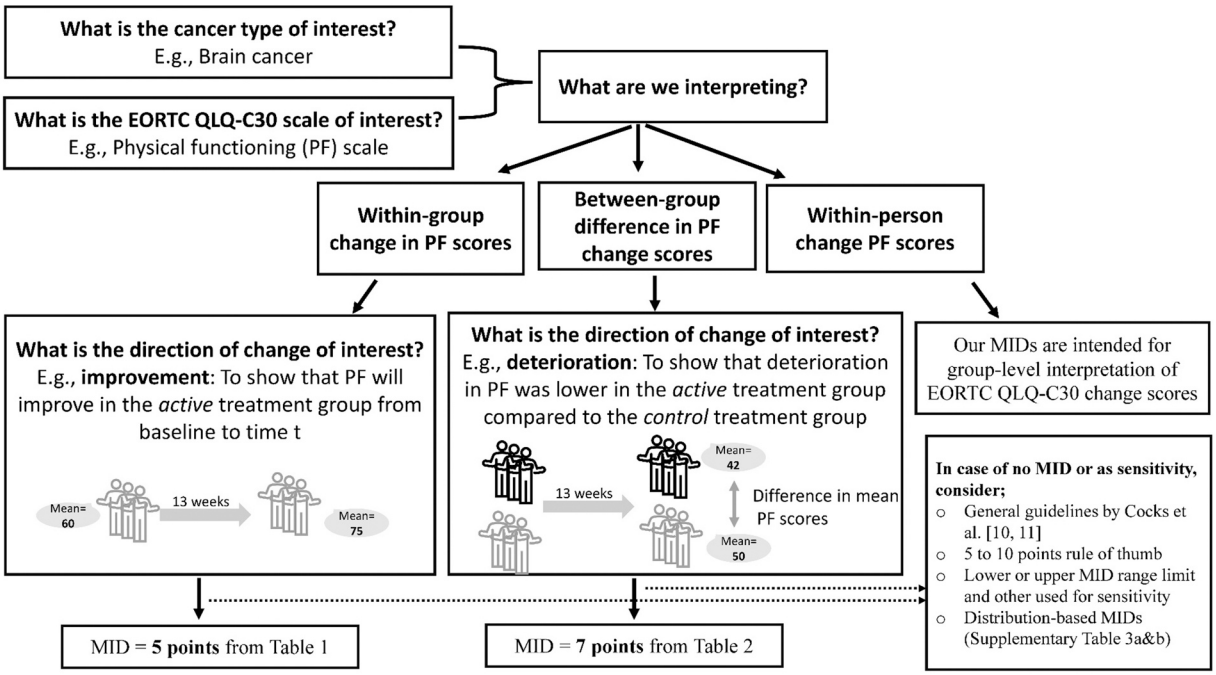

Fig. 3. A flowchart on how to select Minimally Important Differences (MIDs) in practice. *Consider the general MID guidelines for the EORTC QLQ-C30 by Cocks et al. [10,11], 5–10 points rule of thumb (Osoba et al. [9]), lower or upper MID range limit (and other used for sensitivity), or distribution-based estimates (Supplementary Table 3a & b).

distribution-based criteria and multiple HRQoL scales. To aid interpretation of EORTC QLQ-C30 results in clinical research, we have provided both MID ranges per scale, as well as single value MIDs by calculating a correlation-weighted average across multiple anchors (Tables 1 and 2). Fig. 3 provides a flowchart on how to select the single-value MIDs. When ranges are used, worthwhile treatment effect(s) may be defined along this range. One of the limits could be set as the threshold of interest and the other used for sensitivity. The upper limit could be targeted in settings where relatively large changes in HRQoL scores are required to claim benefit. For instance, a more demanding treatment should result in a relatively large HRQOL effect to offset the side-effects. On the other hand, the lower limit could be targeted as an acceptable difference when comparing treatments in a non-inferiority setting. Ultimately, when selecting MIDs, it is crucial to carefully consider the specific settings (e.g. cancer type, QLQ-C30 scale of interest, within/between-group comparison and direction of change) and clinical decision context. Although MIDs tended to vary in our study by aforementioned factors, they mostly ranged from 5 to 10 points, and did not depend on confounders such as age and gender (except for breast, prostate, and ovarian cancer; data not shown). While these results supports previous guidelines [9] and may be easier to apply in practice as it aligns with the commonly used 10-points rule, end-users should still be aware of the risk of dismissing changes that are clinically meaningful or underpowering analyses for scales when smaller thresholds apply. For most scales, especially in ovarian and colorectal cancer, smaller MIDs were observed for deteriorations compared to improvements. One possible explanation for this finding could be prospect theory [32], It will be interesting to further investigate this observation in future research.

Our MIDs are intended for group-level interpretation of QLQ-C30 change scores only. Although it is tempting, our results cannot be directly used for defining within-patient thresholds that are clinically meaningful, due to two caveats. First, since every QLQ-C30 scale has a limited number of observable values, not all MID values can translate to a change score that is achievable for an individual patient. For example, a patient can only change by 33 points for QLQ-C30 single-item scales, such as diarrhoea, whereas multi-item scales (e.g. physical functioning) have many more intermediate values and hence more continuous change scores [33]. Thus, selection of within-patient thresholds should be done with knowledge of the underlying score distribution. Secondly, individual thresholds must be set above limits of measurement error, for example, threshold for a given QLQ-C30 scale should be above their respective SEM estimate (Supplementary Tables 3a and 3b) to avoid false positive changes that may trigger clinical actions [34]. A recent study has published thresholds for interpreting within-patient changes on the QLQ-C30 in patients with non-small cell lung cancer [35]. Giesinger et al. also published thresholds to aid the interpretation of patient-level QLQ-C30 scale scores that are observed during single (cross-sectional) visits in clinical practice [36].

Our research has some limitations. The data used for this study were derived from randomised controlled trials (RCTs), each with specific eligibility criteria, treatment interventions, and outcome assessments, which may limit the generalisability of our results beyond RCTs. The selection of cancer types and clinical anchors was limited to data available in the various EORTC study databases. This makes it difficult to verify whether variability in the MID estimates is due to sample variability or due to genuine differences in the underlying construct represented by the anchor and true variability in MIDs by different cancer types. Although multiple anchors per scale were targeted, for some scales across cancer types only one suitable anchor was found and for other scales no suitable anchor was available, precluding the calculation of MID estimates. Furthermore, the identified anchors were mainly WHO PS and CTCAE grades and were not necessarily suitable in all situations. The correlations between change scores of QLQ-C30 scales and anchors were sometimes suboptimal, that is, less than the recommended 0.3 threshold [26]. The low correlations could be attributed to the subjective nature of the clinical anchors, which were based on physicians' judgements that may deviate from patient-reported scores [31]. In our study, we assumed that a one-category change in each anchor represented a 'minimal' relevant change. However, this assumption may not always hold, and changes or differences below our calculated MIDs may not necessarily be considered trivial. MIDs for interpreting within-group change (presented in Table 1) were estimated by the mean change in QLQ-C30 scores of patients who minimally improved or deteriorated on the clinical anchor, respectively. A variation of this approach has been applied where the MIDs for improvement and deterioration are re-calibrated by subtracting the mean change score of the stable group from mean of the minimally improved and deteriorated groups, respectively [37]. In our study, the ES for the stable group were mostly small or negligible, and we have previously published details on this by cancer type [13–20]. Our estimated MIDs and corresponding confidence intervals (CI) were based on the available data in our study database. No initial sample size calculations were performed to determine the appropriate sample size for estimating MIDs [13–20]. The observed differences in the width of the CIs in our study may have been purely due to the varying sample size by cancer type. Given this limitation, we refrained from making recommendations on MID selection or power calculations based on the CIs as not to over-interpret the results. Anchors that are based on patients' perspective of change such as global ratings of change were not collected in the trials included in this study. We are therefore embarking on a new project that seeks to establish MIDs for all QLQ-C30 scales using an anchor that is based on patient-reported ratings of change over time [9]. This will entail prospective data collection, encompassing multiple cancer types, different disease stages, and treatment settings. Results from this prospective project will contribute to validating the current clinical anchor-based MIDs and will also aid to inform the further refinement of the current QLQ-C30 interpretation guidelines. Despite the limitations, it is reassuring that our MIDs were mostly consistent with other existing guidelines, notwithstanding the differences in the methodological approach, anchor type or patient population [8–11, 24–26].

In conclusion, our findings supplement existing work to build more robust MID guidelines for group-level interpretation of QLQ-C30 change scores. Consistent with recent guidelines, these results reinforce the fact that no single MID can be applied for all QLQ-C30 scales and across various disease conditions. Hence, simple rules of thumb, should be applied with caution. We present a diverse range of MIDs to inform more accurate sample size calculations for clinical trials with EORTC QLQ-C30 end-points.

Ethical approval

Not applicable. This research project was reviewed internally by EORTC Headquarters. The use of the patient data from the various studies fell under their original informed consent wording. Therefore, no additional patient consent was needed from patients or local ethical committees. The original studies were conducted in compliance with the Declaration of Helsinki.

Funding

This study was funded by a grant from the EORTC Quality of Life Group (Grant Number: 006/2014).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the members of the EORTC Melanoma, Breast cancer, Head and Neck cancer, Genito-urinary cancer, Gynecological cancer, Gastro-intestinal cancer, Brain cancer, Lung cancer, and Quality of Life Groups and their clinical investigators, as well as all patients who participated in the analysed trials.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejca.2023.04.027.

References

[1] Kluetz PG, O'Connor DJ, Soltys K. Incorporating the patient experience into regulatory decision making in the USA, Europe, and Canada. Lancet Oncol 2018;19:e267–74.

[2] Bottomley A, Flechtner H, Efficace F, et al. Health related quality of life outcomes in cancer clinical trials. Eur J Cancer 2005;41:1697–709.

[3] Calvert M, Kyte D, Mercieca-Bebber R, Slade A, Chan AW, King MT. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the SPIRIT-PRO extension. JAMA 2018;319:483–94.

[4] Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. JAMA 2013;309:814–22.

[5] Coens C, Pe M, Dueck AC, et al. International standards for the analysis of quality-of-life and patient-reported outcome endpoints in cancer randomised controlled trials: recommendations of the SISAQOL Consortium. Lancet Oncol 2020;21(2):e83–96.

[6] Guyatt GH, Osaba D, Wu AW, et al. Methods to explain the clinical significance of health status measures. Mayo Clinic 2002;77(4):371–83.

[7] Giesinger JM, Efficace F, Aaronson N, et al. Past and current practice of patient-reported outcome measurement in randomized cancer clinical trials: a systematic review. Value Health 2021;24(4):585–91. https://doi.org/10.1016/j.jval.2020.11.004.

[8] King MT. The interpretation of scores from the quality of life questionnaire QLQ-C30. Qual Life Res 1996;5(6):555–67.

[9] 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):139–44.

[10] Cocks K, King MT, Velikova G, et al. Quality, interpretation and presentation of European Organisation for Research and Treatment of Cancer quality of life questionnaire core 30 data in randomised controlled trials. Eur J Cancer 2008;44:1793–8.

[11] 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;48:1713–21.

[12] Musoro ZJ, Hamel J-F, Ediebah DE, et al. Establishing anchor-based minimally important differences (MID) with the EORTC quality of life measures: a meta-analysis protocol. BMJ Open 2017;7:e019117https://doi.org/10.1136/bmjopen-2017-019117.

[13] Dirven L, Musoro JZ, Coens C, et al. Establishing anchor-based minimally important differences for the EORTC QLQ-C30 in glioma patients. Neuro Oncol 2021;23(8):1327–36. https://doi.org/10.1093/neuonc/noab037. PMID: 33598685; PMCID: PMC8328025.

[14] Musoro JZ, Coens C, Singer S, et al. Minimally important differences for interpreting European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 scores in patients with head and neck cancer. Head Neck 2020;42(11):3141–52. https://doi.org/10.1002/hed.26363. Epub 2020 Jul 6. PMID: 32627261.

[15] Musoro JZ, Sodergren SC, Coens C, et al. Minimally important differences for interpreting the EORTC QLQ-C30 in patients with advanced colorectal cancer treated with chemotherapy. Colorectal Dis 2020;22(12):2278–87. https://doi.org/10.1111/codi.15295.

[16] Musoro JZ, Coens C, Greimel E, et al. Minimally important differences for interpreting European Organisation for Research and Treatment of Cancer (EORTC) Quality of life Questionnaire Core 30 scores in patients with ovarian cancer. Gynecol Oncol 2020;159(2):515–21. https://doi.org/10.1016/j.ygyno.2020.09.007. Epub 2020 Sep 21. PMID: 32972782.

[17] Musoro ZJ, Bottomley A, Coens C, et al. 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;104:169–81. https://doi.org/10.1016/j.ejca.2018.09.005.

[18] Musoro JZ, Coens C, Fiteni F, et al. Minimally important differences for interpreting European Organisation for Research and Treatment of Cancer (EORTC) Quality of life Questionnaire core 30 scores in patients with advanced breast cancer. JNCI Cancer Spectrum. pkz037, https://doi.org/10.1093/jncics/pkz03.

[19] Gamper EM, Musoro JZ, Coens C, et al. Minimally important differences for the EORTC QLQ-C30 in prostate cancer clinical trials. BMC Cancer 2021;21:1083. https://doi.org/10.1186/s12885-021-08609-7.

[20] Koller M, Musoro JZ, Tomaszewski K, et al. Minimally important differences of EORTC QLQ-C30 scales in patients with lung cancer or malignant pleural mesothelioma – Interpretation guidance derived from two randomized EORTC trials. Lung Cancer, 2022, ISSN 0169-5002, https://doi.org/10.1016/j.lungcan.2022.03.018.

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

[22] 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.

[23] Harper A, Trennery C, Sully K, Trigg A. Triangulating estimates of meaningful change or difference in patient-reported outcomes: application of a correlation-based weighting procedure. Paper presented at: Quality of Life Research. 2018.

[24] 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):1249–54.

[25] King MT. A point of minimal important difference (MID): a critique of terminology and methods. Expert Rev Pharmacoecon Outcomes Res 2011;11(2):171–84.

[26] Cohen J. Statistical power analysis for the behavioural sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

[27] 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.

[28] Maringwa JT, Quinten C, King MT, et al. on behalf of the EORTC PROBE project and the Lung Cancer Group. 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. Support Care Cancer 2011;19(11):1753–60.

[29] Maringwa J, Quinten C, King M, et al. Minimal clinically meaningful differences for the EORTC QLQ-C30 and EORTC QLQ-BN20 scales in brain cancer patients. Ann Oncol 2011;22(9):2107–12.

[30] Kawahara T, Taira N, Shiroiwa T, et al. Minimal important differences of EORTC QLQ-C30 for metastatic breast cancer patients: results from a randomized clinical trial. Qual Life Res 2022;31(6):1829–36. https://doi.org/10.1007/s11136-021-03074-y. Epub 2022 Jan 4. PMID: 34982354; PMCID: PMC9098551.

[31] Basch E, Dueck AC, Rogak LJ, et al. Feasibility assessment of patient reporting of symptomatic adverse events in multicenter cancer clinical trials. JAMA Oncol 2017;3(8):1043–50.

[32] Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk (PDF). Econometrica 1979;47(2):263–91.
[33] Cocks K, Buchanan J. How scoring limits the usability of minimal important differences (MIDs) as responder definition (RD): an exemplary demonstration using EORTC QLQ-C30 subscales. Qual Life Res 2023;32(5):1247–53. https://doi.org/10.1007/s11136-022-03181-4. Epub 2022 Jul 9. PMID: 35809136.
[34] King MT, Dueck AC, Revicki DA. Can methods developed for interpreting group-level patient-reported outcome data be applied to individual patient management? Med Care 2019;57(Supp. 1):S38–45.
[35] Coon CD, Schlichting M, Zhang X. Interpreting within-patient changes on the EORTC QLQ-C30 and EORTC QLQ-LC13. Patient 2022. https://doi.org/10.1007/s40271-022-00584-w.
[36] Giesinger JM, Kuijpers W, Young T, et al. Thresholds for clinical importance for four key domains of the EORTC QLQ-C30: physical functioning, emotional functioning, fatigue and pain. Health Qual Life Outcomes 2016;14:87. https://doi.org/10.1186/s12955-016-0489-4.
[37] Angst F, Aeschlimann A, Angst J. The minimally clinically important difference raised the significance of outcome effecs above the statistical level, with methodological implications for future studies. J Clin Epidemiol 2017;82:128–36.