Prepared for: *Lung Cancer*

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

Michael Koller[1]*, Jammbe Z Musoro[2]*, Krzysztof Tomaszewski[3] Corneel Coens[2], Madeleine T King[4], Mirjam AG Sprangers[5], Mogens Groenvold[6], Kim Cocks[7], Galina Velikova[8], Hans-Henning Flechtner[9] Andrew Bottomley[2] on behalf of the EORTC Quality of Life and Gastro-intestinal Groups

[1]Center for Clinical Studies, University Hospital Regensburg, Regensburg, Germany.
[2]European Organisation for Research and Treatment of Cancer (EORTC), Brussels, Belgium
[3] Faculty of Medicine and Health Sciences, Andrzej Frycz Modrzewski Kraków University, Kraków, Poland.
[4]University of Sydney, Faculty of Science, School of Psychology, Sydney, NSW, Australia
[5]Department of Medical Psychology, Amsterdam University Medical Center, Amsterdam, The Netherlands.
[6]Department of Public Health, University of Copenhagen, and Bispebjerg Hospital, Copenhagen, Denmark
[7]Adelphi Values, Bollington, Cheshire, UK
[8]Leeds Institute of Medical Research at St James's, University of Leeds, Leeds, UK.
[9]Clinic for Child and Adolescent Psychiatry and Psychotherapy, University of Magdeburg, Magdeburg, Germany

Corresponding Author:
Dr Michael Koller
Center for Clinical Studies, University Hospital Regensburg, Regensburg, Germany
Michael.Koller@ukr.de

*Joint first authorship

ABSTRACT

Objectives

A minimally important difference (MID) is the smallest difference in quality of life (QoL) perceived as relevant by patients or clinicians. MIDs aid interpretation of QOL data in research and clinical practice. We aimed to determine MIDs for the EORTC QLQ-C30 for patients with lung cancer or malignant pleural mesothelioma.

Materials and Methods

Data were drawn from two EORTC-sponsored randomised clinical trials (RCTs): a three-arm RCT of two cisplatin-based treatments and paclitaxel plus gemcitabine in advanced non–small-cell lung cancer, and an RCT comparing cisplatin with or without raltitrexed in patients with malignant pleural mesothelioma. MIDs for interpreting within-group change and between-group differences in change over time were computed using anchor-based approaches, for improvements and deteriorations separately. Distribution-based approaches provided corroborative evidence.

Results

The combined data from the trials comprised 730 patients. Available data allowed us to determine 8/14 anchor-based MIDs for EORTC scales for improvements, and 9/14 MIDs for deterioration. Furthermore, we provided distribution-based estimates for all 14 QLQ-C30 scales. Most MIDs for improvements ranged between 5 and 10, for both within-group and between-group differences. Outliers were appetite loss and constipation, with MIDs up to 15 score points. MIDs were slightly larger for within-group deterioration, ranging from –5 to -15, with the largest for Nausea/vomiting (-14) and Appetite loss (-15). MIDs for between-group differences in deterioration ranged from -4 (Physical, Role, and Social functioning, and Global quality of life) to -9 (Nausea/vomiting, Appetite loss and Constipation).

Conclusions

MIDs vary over scales and for between- versus within-group comparisons; this must be taken into account when interpreting changes. Nevertheless, the majority of MIDs range between 5 and 10 score points, in line with previously used thresholds for QLQ-C30. These findings and those from other tumor-specific MID analyses will inform a planned consensus process identifying commonalities and differences across tumor sites.

Keywords: Lung cancer, pleural mesothelioma, minimally important difference (MID), clinical anchors, quality of life (QoL), EORTC QLQ-C30

INTRODUCTION

Lung cancer is the most prevalent cancer worldwide and accounts for the highest number of cancer deaths [1]. Smoking is the number one cause of lung cancer [2,3,4,5], but also genetics [6,7], pollution and occupational exposure [8, 9, 10] , socioeconomic factors [11], as well as gender [12, 13] play a role in its epidemiology. Prognosis is poor, although the 5-year relative survival rates for all types of lung cancer increased from 11% to 20% in the past four decades [14]. This increase is due to in part to earlier detection but moreso to progress in systemic therapies such as targeted therapy and immunotherapy [15, 16, 17].

There are two main types, Non-Small Cell Lung Cancer (NSCLC), accounting for approximately 85% of cases, and Small Cell Lung Cancer (SCLC). Treatment decisions are based on cancer type, stage, histology and grade, performance status, and comorbid conditions. Treatment options for lung cancer are surgery, radiotherapy, chemotherapy, and advanced systemic therapies including targeted therapy and immunotherapy. In an attempt to improve clinical outcomes, these treatment options are often administered in a multimodality approach [18, 19].

Another malignancy that differs from lung cancer in aetiology but is associated with similar somatic symptoms (dyspnea, pain in the chest, fatigue) is malignant pleural mesothelioma (MPM). Compared to lung cancer, MPM is a relatively rare disease with worldwide incidences rates per 100.000 ranging from 0,6 (Poland) to 3,4 (United Kingdom) [20]. The major cause of MPM is exposure to asbestos (90%), and most affected patients are men (80%) [21]. The outcome is unfavourable, with median survival less than 1 year [22]. Treatment options include chemotherapy, radiotherapy and surgery, often administered in a multimodal strategy [21, 22]. Despite the different nature and trajectory of the diseases, lung cancer and MPM have common respiratory problems and impact on patients' activities and QoL, which justifies combining the two samples for the present purposes.

Malignant conditions and their treatments often affect patients' quality of life (QoL). Therefore, QoL is now assessed as an endpoint in most cancer clinical trials. A commonly used QoL assessment tool in cancer clinical trials is the European Organisation of Research and Treatment of Cancer (EORTC) QLQ-C30 [23], which covers core aspects of quality of life of cancer patients, including five aspects of functioning and eight common symptoms. It is supplemented by a lung-cancer specific module, QLQ-LC13 [24], which was recently updated in response to the improvements in diagnostics and therapy mentioned above (QLQ-LC29) [25]. The EORTC

QLQ-C30 and the QLQ-LC13 have also been successfully used and validated in patients with MPM [26].

When it comes to interpreting changes or differences in QoL, the concept of minimal important difference (MID) is crucial. MID is commonly defined as the smallest change in QoL that is perceived as "relevant" or "important" by a patient or by a third party (eg, physician), and which may affect the course of patient management [27].

In the context of a larger project, the EORTC is exploring MIDs of its core questionnaire QLQ-C30 in a series of studies involving a variety of diagnoses [28, 29, 30, 31, 32, 33, 34]. The focus of the present paper is lung cancer and malignant pleural mesothelioma, and is based on data collected in two EORTC-sponsored clinical trials. The goal of this secondary analysis is twofold: to determine MIDs for interpreting (1) within-group change in EORTC QLQ-C30 scores over time and (2) between-group differences in EORTC QLQ-C30 change scores over time.

METHODS

Description of the data

The data were pooled across two published EORTC phase III trials. The first trial (EORTC 08975) is a three-arm randomized study of paclitaxel/cisplatin, gemcitabine/cisplatin and paclitaxel/gemcitabine in advanced non–small-cell lung cancer; it enrolled 480 patients [35]. The second trial (EORTC 08983) compared cisplatin with or without raltitrexed in patients with malignant pleural mesothelioma; it enrolled 250 patients [22]. Both trials assessed QoL at baseline and multiple time points during treatment, using the EORTC QLQ-C30. Both trials were approved by the ethical committee of each participating center, and all patients consented to participate in their respective studies. Both trials were registered with clinicaltrials.gov, NCT00003589 and NCT00004920, respectively.

The EORTC QLQ-C30

This instrument comprises five functional scales (physical, role, emotional, social, and cognitive), a global health status/QoL scale, three multi-item symptom scales (fatigue, nausea and vomiting, and pain), and six single items (dyspnea, insomnia, appetite loss, constipation, diarrhea and financial difficulties). All scales meet the standards for reliability. The reliability and validity of the questionnaire has been highly consistent across different language-cultural

groups [36]. Both trials used version 3 of the EORTC QLQ-C30, and individual item responses were scored according to the scoring manual [36], so that scales and single item scores ranged from 0 to 100. For this paper, we deviated from standard scoring in one way to achieve consistency in scale direction: all scales were scored such that 0 represents the worst and 100 the best possible score. Thus, in this paper, improvement is always represented by a positive sign and deterioration is always represented by a negative sign, for all scales and items. The financial difficulties scale was excluded from the analysis since suitable anchors were unavailable.

Clinical anchors

Anchors for each EORTC QLQ-C30 scale were selected from available clinical variables such as specific common terminology criteria for adverse events (CTCAE), physician examinations and WHO performance status (PS). Anchors for each scale were selected based on the correlation strength and clinical plausibility. Priority was given to anchor and scale pairs with correlations of $\geq|0.30|$, as recommended for determining MIDs [37]. Where attainable, anchors with stronger correlations were preferred as proposed by simulation studies [38]. For scales where the 0.30 threshold was unachievable for the majority of the anchors, we selected anchors with a mixture of weak (<0.30) to optimal correlations. Where available, multiple anchors per scale were used to offer assurance about the reliability of the estimated MIDs. The retained anchors were further screened for clinical plausibility by four clinical and HRQOL experts to avoid unreasonable findings. Details on the procedure for selecting anchors have earlier been described in a published protocol [39].

Definition of clinical anchor change groups

For the retained anchors per EORTC QLQ-C30 scale, the following clinical change groups were defined according to anchor change scores: *(i)* deterioration; worsened by 1 anchor category, *(ii)* stable; no change in anchor category, and *(iii)* improvement; improved by 1 anchor category. Change scores ≥ 2 points in anchor categories were excluded from data used for MID estimation since they were considered to be obviously above the "minimal" expected change.

Data analysis

We applied anchor- and distribution-based methods to the pooled data set of QoL scales/items, following methods used in our previous MID studies [28, 29, 30, 31, 32, 33]. As these methods are described in detail in our earlier papers, we describe briefly below. All statistical analyses were performed using the SAS software [40].

Anchor-based methods

The first anchor-based method estimated *within-group MIDs*, separately for improvement and deterioration, by calculating the mean change QoL score of patients within the improvement and deterioration clinical change groups respectively. This approach is known as the mean-change method and it is useful for interpreting change in QoL scores over time within a group of patients.

The second anchor-based method applied a linear regression model to *estimate between-group MIDs* for interpreting differences in changes over time between two distinct groups of patients. For each QoL scale/item and anchor pair, a regression model was fitted with the QoL change score as the outcome and a binary anchor factor coded as 'stable' = 0 and 'improvement' = 1 when modelling improvement (excluding deteriorated observations) and 'stable' = 0 and 'deterioration' = 1 when modelling deterioration (excluding improved observations). The resulting slope parameters for the 'improved' and 'deteriorated' covariates are the estimated MIDs for improvement and deterioration respectively.

When multiple anchor-based MID estimates were available for the same QoL scale/item, they were triangulated to a single value by calculating a weighted estimate using the correlations between anchor and QoL change scores as weights. This was done separately for within-group MIDS and between-group MIDS. Also, to check if MIDs depend on trial (i.e., lung cancer versus pleural mesothelioma), a binary trial factor and its interaction with the binary anchor factor were added to the regression models (separately for improving and deteriorating QoL scores); this was possible for between-group MIDS only. To account for multiple testing across the various scales/items and direction of change, p-values below 0.001 were considered to be statistically significant in our regression models.

Distribution-based methods

The distribution-based estimates considered included 0.2 SD, 0.3 SD and 0.5 SD and standard error of measurement (SEM) [37]. These were calculated from the pooled data for each scale/item using the time point before or on the first day of treatment administration (t1).

In addition to these cross-sectional estimates, within-group effect sizes (ES) were computed within every clinical change group by taking the mean of the QoL change scores divided by the SD of the change scores. These ES were classified according to Cohen's [41] guidelines, i.e. an

ES of 0.2 is small, 0.5 is moderate and ≥0.8 is large. Hence, only mean changes with effect sizes ≥ 0.2 and <0.8 were considered appropriate for inclusion as MID estimates, as an observed ES <0.2 reflects clinically unimportant changes and ESs ≥0.8 are obviously more than minimally important.

RESULTS

The combined data from the trials comprised 730 patients. A summary of patient and disease characteristics at baseline are presented in Table 1. The mean follow-up time for QoL was around 5 months and 4 months for the lung and MPM trial respectively.

A total of 17 potential clinical anchors were initially evaluated for each EORTC QLQ-C30 scale, based on available clinical variables in both trials. After prioritising anchors with at least a 0.3 cross-sectional correlation and expert endorsement of their clinical relevance, 9 anchors were finally retained. One such anchor was WHO performance status (PS), scored between 0 (no cancer symptoms) and 4 (bedbound). The other 8 anchors referred to various CTCAEs (pain, fatigue, nausea & vomiting, diarrhoea, flulike syndrome, anorexia, gastrointestinal symptoms and constipation), all graded between 0 (no toxicity) and 4 (life-threatening).

At least one anchor was identified for 11 of the 14 EORTC QLQ-C30 scales evaluated. In Table 2, cross-sectional correlations between QoL scale/item scores and anchors ranged from 0.33 to 0.66 in absolute values, while correlations between their change scores where in the range of 0.20 to 0.50. The distribution of patients (and number of change observations) by the anchor change categories are summarised in Table A.1. The majority of the patients remained stable compared to those who improved or deteriorated.

A summary of the anchor-based MIDs for interpreting change in QoL over time within and between-groups of patients is presented in Table 3. This considered only estimates from anchor clinical change groups with an ES of ≥ 0.2 and <0.8. The complete results are presented in Table A.2. Anchor-based MIDs were determined for improving scores in 8 of the 14 EORTC QLQ-C30 scales assessed, and in 9 scales for deteriorating scores. A weighted MID average is also presented for scales with multiple MIDs from different anchors (Table 3).

In general, anchor-based MIDs varied somewhat by scale, anchor and, direction of change (improvement versus deterioration), as illustrated in Figure 1 where within-group MIDs from the mean change method (from Table 3) are plotted along with their 95% confidence intervals. As expected, positive change scores were observed within the improvement clinical change group and negative change scores within the deterioration clinical change group. As shown in Table 3, within-group MIDs ranged from 5 to 13 points for improvement and -15 to -5 points

for deterioration, whereas between-group MIDs for improvement ranged from 4 to 15 points and deterioration -9 to -4. The MID estimates did not differ significantly by trial (results not shown). Overall, the MID estimates for most EORTC QLQ-C30 scales were in the range of 4 to 10 points in absolute values for both within and between-group change.

Distribution-based MIDs for the 14 assessed EORTC QLQ-C30 scales are presented in Table A.3. Compared to the distribution-based estimates, anchor-based MIDs for improvement were mostly in the range of 0.2 SD and 0.3 SD, except for Constipation with estimates closer to the 0.5 SD and 1 SEM. Anchor-based MIDs for deterioration were mainly in the range of 0.2 SD and 0.5 SD.

DISCUSSION

This is the most comprehensive analysis to date to specify MIDs for within-group comparisons as well as between-group comparisons of QLQ-C30 change scores over time for lung cancer and MPM patients. Available data allowed us to determine 8/14 anchor-based MIDs for QLQ-C30 scales for improvements, and 9/14 MIDs for deterioration. Furthermore, we provided distribution-based estimates for 14 QLQ-C30 scales; this information backs anchor-based findings and supports interpretations when no anchor-based MIDs were available.

Most anchor-based MIDs for improvements ranged between 5 and 10 score points, for both within-group and between-group differences. Outliers were appetite loss and constipation with MIDs up to 15 score points. MIDs were slightly larger for within-group deterioration, ranging from –5 to -15, with the largest for Nausea/vomiting (-14) and Appetite loss (-15). MIDs for between-group differences in deterioration ranged from -4 (Physical, Role, and Social functioning, and Global quality of life) to -9 (Nausea/vomiting, Appetite loss and Constipation). These MID estimates apply to lung cancer and MPM trials, as we found no statistically significant differences in MID estimates between the two trials used. This suggests that we may apply the derived MIDs with relative confidence to more recent trials on immunotherapy and targeted therapy for lung cancer.

Maringwa et al. (2011) [42] had earlier published MIDs for 6 scales of the QLQ-C30 in lung cancer. Similar to this current study and previously published guidelines [43,44,45,46], our previous MID estimates were mostly in the 5 to 10 points range. In contrast to Maringwa et al., (2011), one of the trials used in the current analyses included data of patients with malignant pleural mesothelioma, whereas both trials used in the 2011 analyses included patients with non–small-cell lung cancer. Also, our study pooled data on change scores that were computed across all possible pairwise timepoints [28, 29, 30, 31, 32, 33] while in 2011, we only looked at change between the two furthest two times points. Finally, our derived MIDs were based on 8 clinical anchors whereas in 2011 we used only two anchors, with relatively slightly weaker anchor versus QoL scale/item correlations.

Our current study is the latest in a series of post-hoc analyses taking advantage of published EORTC trials on brain cancer, breast cancer, colorectal cancer, head and neck cancer, ovarian cancer, prostate cancer, and melanoma [28, 29, 30, 31, 32, 33, 34]. By and large the present data echo the pattern of the results of the previous studies.

There are at least two ways to look at these results: on a granular level, certain differences were obtained in the size of anchor-based MIDs for the various QoL scales/items, and whether improvement or deterioration as well as within-group or between group differences were at the core. These differences, although not large, may be relevant for designing future clinical studies

that intend to focus on a particular QoL scale/item as a primary or co-primary endpoint and wish to explore this endpoint in a pre-specified context (improvement or deterioration; within-group or between-group change).

Adopting an overarching perspective, however, the present study as well as the other papers in this series of analyses, reveal that a large number of MIDs fall within the range of 5 to 10 score points, which is reminiscent of Osoba's initial suggestion [43]. On a superficial level this can be seen to support the commonly used 10 points rule of thumb which, however, would risk underpowered analyses for many domains if used to determine sample size and would risk discounting clinically important findings less than 10 points. Growing empirical evidence that MIDs vary by scale, direction of change and by selected anchor casts doubt on the use of on global standards for defining clinically meaningful change. As proposed by Cocks et al [45], and supported by our series of studies [28, 29, 30, 31, 32, 33, 34], the solution lies in guidelines that advocate for a more nuanced, yet manageable, strategy to clinical relevance beyond a single threshold.

The present results are limited since they are confined to two EORTC trials and involve moderate sample sizes. Furthermore, although these were international multicenter trials, the majority of the included patients stem from the Netherlands. Thus, clearly the generalizability of the present findings would profit from further analyses using other data sets outside the EORTC context, with larger samples and possibly with different anchors. Furthermore, we relied on anchors that were derived from data collected in two randomized clinical trials for safety purposes. Clinically appropriate anchors could not found for all QLQ-C30 scales from available data, Therefore, we were not able to specify MIDs for 7 respectively 6 of the total of 15 QLQ-C30 scales. However, note that there are a number of published articles that provide general guidance on MID selection for all 15 QLQ-C30 scales [44, 45]. Our estimates generally aligned with the guidelines for small differences that was proposed by Cocks *et al* [44] for interpreting changes in QLQ-C30 scores over time. Also, most of the correlations between change scores of anchors and QoL scales were suboptimal (<0.30 but >=0.20). However, applying a mixture of anchors (with varying correlation strengths) for several QoL scales/items provided some reassurance about the plausibility of the MID estimates. Furthermore, the present analyses were based on data sets that were originally published in 2003 and 2005 and did not include immunotherapy. This may be seen as a limitation, but also as in interesting extension for future analyses.

Conclusion

MIDs vary over scores/items, anchors and between versus within group comparisons. This should be taken into when interpreting changes. Nevertheless, the majority of MIDs range between 5 and 10 score points which is in line with the commonly used rule of thumb of interpreting QLQ-C30 change scores. An overarching consensus statement taking into account the findings from the various tumor-site specific MID analyses is warranted to specify when to use granular scale-specific MIDs and when a rule-of-thumb approach is sufficient.

Acknowledgements: We thank the members of the EORTC Lung cancer group and their clinical investigators, as well as all patients who participated in the analyzed trials.

Funding: This study was funded by an academic grant from the EORTC Quality of Life Group (Grant Number: EORTC MID Grant).

Competing interests: The authors report no conflicts of interest in this work.

Ethical approval: Not applicable**.** This research project was checked internally by EORTC HQ. The use of the patient data from the various studies fell under their original informed consent wording. So, no additional patient consent was needed from patient or local ethical committees. The original studies were conducted in compliance with the Declaration of Helsinki.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; 68: 394–424.
2. Islami F, Torre LA, Jemal A· Global trends of lung cancer mortality and smoking prevalence· Translational lung cancer research 2015; 4: 327–338·
3. Samet JM· Is there more to learn about the epidemiology of lung cancer? European journal of epidemiology 2016; 31: 1159–1160·
4. Tanner NT, Kanodra NM, Gebregziabher M, Payne E, Halbert CH, Warren GW, Egede LE et al· The Association between Smoking Abstinence and Mortality in the National Lung Screening Trial· American journal of respiratory and critical care medicine 2016; 193: 534–541·
5. Alberg AJ, Shopland DR, Cummings KM· The 2014 Surgeon General's report: Commemorating the 50th Anniversary of the 1964 Report of the Advisory Committee to the US Surgeon General and updating the evidence on the health consequences of cigarette smoking· American journal of epidemiology 2014; 179: 403–412·
6. Sholl LM· The Molecular Pathology of Lung Cancer· Surgical pathology clinics 2016; 9: 353–378·
7. Bossé Y, Amos CI· A Decade of GWAS Results in Lung Cancer· Cancer epidemiology, biomarkers & prevention a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2018; 27: 363–379·
8. Cui P, Huang Y, Han J, Song F, Chen K· Ambient particulate matter and lung cancer incidence and mortality: A meta-analysis of prospective studies· European journal of public health 2015; 25: 324–329·
9. Hamra GB, Guha N, Cohen A, Laden F, Raaschou-Nielsen O, Samet JM, Vineis P et al· Outdoor particulate matter exposure and lung cancer: A systematic review and meta-analysis· Environmental health perspectives 2014; 122: 906–911·
10. Santibáñez-Andrade M, Quezada-Maldonado EM, Osornio-Vargas Á, Sánchez-Pérez Y, García-Cuellar CM· Air pollution and genomic instability: The role of particulate matter in lung carcinogenesis· Environmental pollution (Barking, Essex 1987) 2017; 229: 412–422·
11. Hovanec J, Siemiatycki J, Conway DI, Olsson A, Stücker I, Guida F, Jöckel K-H et al· Lung cancer and socioeconomic status in a pooled analysis of case-control studies· PloS one 2018; 13: e0192999·

12. Welcker K· Genderspezifische Unterschiede des Lungenkarzinoms· Zentralblatt für Chirurgie 2015; 140: 260–265·
13. Isla D, Majem M, Viñolas N, Artal A, Blasco A, Felip E, Garrido P et al· A consensus statement on the gender perspective in lung cancer· Clinical & translational oncology official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico 2017; 19: 527–535·
14. Lu T, Yang X, Huang Y, et al. Trends in the incidence, treatment, and survival of patients with lung cancer in the last four decades. Cancer Manag Res 2019; 11: 943–53.
15. Besse B, Adjei A, Baas P, Meldgaard P, Nicolson M, Paz-Ares L, Reck M et al· 2nd ESMO Consensus Conference on Lung Cancer: Non-small-cell lung cancer first-line/second and further lines of treatment in advanced disease· Annals of oncology official journal of the European Society for Medical Oncology 2014; 25: 1475–1484·
16. Tan W-L, Jain A, Takano A, Newell EW, Iyer NG, Lim W-T, Tan E-H et al· Novel therapeutic targets on the horizon for lung cancer· The Lancet Oncology 2016; 17: e347-e362·1
17. Tartarone A, Giordano P, Lerose R, Rodriquenz MG, Conca R, Aieta M· Progress and challenges in the treatment of small cell lung cancer· Medical oncology (Northwood, London, England) 2017; 34: 110·
18. Eberhardt WEE, Ruysscher D de, Weder W, Le Péchoux C, Leyn P de, Hoffmann H, Westeel V et al· 2nd ESMO Consensus Conference in Lung Cancer: Locally advanced stage III non-small-cell lung cancer· Annals of oncology official journal of the European Society for Medical Oncology 2015; 26: 1573–1588·
19. Jones CM, Brunelli A, Callister ME, Franks KN· Multimodality Treatment of Advanced Non-small Cell Lung Cancer: Where are we with the Evidence? Current surgery reports 2018; 6: 5·
20. Bianchi C, Bianchi T. Global mesothelioma epidemic: Trend and features. Indian J Occup Environ Med. 2014 May;18(2):82-8. doi: 10.4103/0019-5278.146897. PMID: 25568603; PMCID: PMC4280782.
21. Neumann V, Löseke S, Nowak D, Herth FJ, Tannapfel A. Malignant pleural mesothelioma: incidence, etiology, diagnosis, treatment, and occupational health. Dtsch Arztebl Int. 2013 May;110(18):319-26. doi: 10.3238/arztebl.2013.0319. Epub 2013 May 3. PMID: 23720698; PMCID: PMC3659962.
22. van Meerbeeck JP, Gaafar R, Manegold C, Van Klaveren RJ, Van Marck EA, Vincent M, Legrand C, Bottomley A, Debruyne C, Giaccone G; European Organisation for Research and Treatment of Cancer Lung Cancer Group; National Cancer Institute of Canada. Randomized phase III study of cisplatin with or without raltitrexed in patients

with malignant pleural mesothelioma: an intergroup study of the European Organisation for Research and Treatment of Cancer Lung Cancer Group and the National Cancer Institute of Canada. J Clin Oncol. 2005 Oct 1;23(28):6881-9.

23. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A 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· Journal of the National Cancer Institute 1993; 85: 365–376·
24. Bergman B, Aaronson NK, Ahmedzai S, Kaasa S, Sullivan M.The EORTC QLQ-LC13: a modular supplement to the EORTC Core Quality of Life Questionnaire (QLQ-C30) for use in lung cancer clinical trials. Eur J Cancer 1994; 30A: 635–42
25. Koller M, Shamieh O, Hjermstad MJ et al. Psychometric properties of the updated EORTC module for assessing quality of life in patients with lung cancer (QLQ-LC29): an international, observational field study. The Lancet Oncology, Volume 21, Issue 5, 723 – 732
26. Nowak AK, Stockler MR, Byrne MJ. Assessing quality of life during chemotherapy for pleural mesothelioma: feasibility, validity, and results of using the European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire and Lung Cancer Module. J Clin Oncol. 2004 Aug 1;22(15):3172-80. doi: 10.1200/JCO.2004.09.147. PMID: 15284270.
27. Schünemann HJ, Guyatt GH. Goodbye M(C)ID! Hello MID, where do you come from? Health Serv Res. 2005; 40: 593-597.
28. Dirven L, Musoro JZ, Coens C, Reijneveld JC, Taphoorn MJB, Boele FW, Groenvold M, van den Bent MJ, Stupp R, Velikova G, Cocks K, Sprangers MAG, King MT, Flechtner HH, Bottomley A. Establishing anchor-based minimally important differences for the EORTC QLQ-C30 in glioma patients. Neuro Oncol. 2021 Feb 18:noab037. doi: 10.1093/neuonc/noab037. Epub ahead of print. PMID: 33598685.
29. 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 [published online ahead of print, 2020 Jul 6]. Head Neck. 2020;10.1002/hed.26363. doi:10.1002/hed.26363.
30. Musoro, J.Z., Sodergren, S.C., Coens, C., Pochesci, A., Terada, M., King, M.T., Sprangers, M.A.G., Groenvold, M., Cocks, K., Velikova, G., Flechtner, H.-H., Bottomley, A. and (2020), Minimally important differences for interpreting the EORTC QLQ-C30 in patients with advanced colorectal cancer treated with chemotherapy. Colorectal Dis. doi:10.1111/codi.15295.

31. Musoro JM, Coens , 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. Gynecologic Oncology, 2020
32. 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. European Journal of Cancer (2018) 104, 169-181. doi.org/10.1016/j.ejca.2018.09.005
33. Musoro ZJ, 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
34. Gamper, E.M., Musoro, J.Z., Coens, C. et al. Minimally important differences for the EORTC QLQ-C30 in prostate cancer clinical trials. BMC Cancer 21, 1083 (2021). https://doi.org/10.1186/s12885-021-08609-7
35. Smit EF, van Meerbeeck JP, Lianes P, Debruyne C, Legrand C, Schramel F, Smit H, Gaafar R, Biesma B, Manegold C, Neymark N, Giaccone G; European Organization for Research and Treatment of Cancer Lung Cancer Group. 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. 2003 Nov 1;21(21):3909-17. doi: 10.1200/JCO.2003.03.195. PMID: 14581415.
36. Fayers P, Aaronson, NK, Bjordal K, Groenvold M, Curran D, and Bottomley A on behalf of the EORTC Quality of Life Study Group. EORTC QLQ-C30 Scoring Manual (Third edition). Brussels, EORTC Quality of Life Group, 2001.
37. Revicki D, Hays RD, Cella D, Sloan J Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008; 61:102–109
38. Coon CD. Empirical telling the interpretation story: the case for strong anchors and multiple methods. In: 23rd annual conference of the International Society for Quality of Life Research, Copenhagen, Denmark, October 2016; 2016. p. 1e2. Qual Life Res; 25(1), ab2.
39. 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:e019117. doi:10.1136/bmjopen-2017-019117
40. Institute Inc. 2013. Base SAS® 9.4 Procedures Guide. Cary, NC: SAS Institute Inc.

41. Cohen J. Statistical Power Analysis for the Behavioural Sciences (2nd Edition). Lawrence Erlbaum Associates, NJ, USA (1988).
42. Maringwa JT, 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.
43. Osoba D Rodrigues G, Myles J, et al. Interpreting the significance of changes in health related quality-of-life scores. J Clin Oncol. 1998; 16: 139-144.
44. Cocks K, 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. European Journal of Cancer (2012) 48, 1713– 1721.
45. 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) 44:1793-1798.
46. King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. Qual Life Res. 1996; 5: 555-567.

Table 1: Baseline demographic and clinical characteristics of the patients by study

	study		
	EORTC 08975 (N=480)	**EORTC0 8983 (N=250)**	**Total (N=730)**
	N (%)	**N (%)**	**N (%)**
Gender			
Male	318 (66.3)	199 (79.6)	517 (70.8)
Female	162 (33.8)	51 (20.4)	213 (29.2)
Performance status			
0	113 (23.5)	62 (24.8)	175 (24.0)
1	311 (64.8)	155 (62.0)	466 (63.8)
2	56 (11.7)	33 (13.2)	89 (12.2)
Clinical M staging			
M1	388 (80.8)	39 (15.6)	427 (58.5)
M0	92 (19.2)	206 (82.4)	298 (40.8)
Mx	0 (0.0)	5 (2.0)	5 (0.7)
Country			
Netherlands	367 (76.5)	59 (23.6)	426 (58.4)
Egypt	37 (7.7)	43 (17.2)	80 (11.0)
Germany	24 (5.0)	33 (13.2)	57 (7.8)

Table 1: Baseline demographic and clinical characteristics of the patients by study

	study		
	EORTC 08975 (N=480)	**EORTC0 8983 (N=250)**	**Total (N=730)**
	N (%)	**N (%)**	**N (%)**
United Kingdom	12 (2.5)	30 (12.0)	42 (5.8)
Spain	25 (5.2)	0 (0.0)	25 (3.4)
Poland	0 (0.0)	21 (8.4)	21 (2.9)
Canada	0 (0.0)	20 (8.0)	20 (2.7)
France	5 (1.0)	14 (5.6)	19 (2.6)
Belgium	2 (0.4)	10 (4.0)	12 (1.6)
Others[1]	8 (1.7)	20 (8.0)	28 (3.8)
Age			
Mean (SD)	56.23 (9.80)	57.88 (9.34)	56.80 (9.67)
Interquartile	49.0 - 64.0	53.0 - 64.0	51.0 - 64.0

[1]Other countries: Peru, Switzerland, Italy, Czech Republic and South Africa

Table 2: Correlations over all time points of HRQOL scale scores with anchors, and correlations between change scores of the HRQOL scales and change in anchors

		Score		**Change score**	
EORTC QLQ-C30 scale	**Anchor**	**n_{1R} (n_1)***	**Correlation**	**n_{2R} (n_2)***	**Correlation**
Physical functioning	Performance status	2626 (560)	-0.51	5237 (549)	-0.30
	CTCAE Fatigue	2686 (582)	-0.35	5709 (562)	-0.30
	CTCAE Flulike syndrome	2686 (582)	-0.34	5709 (557)	-0.24
Role functioning	Performance status	2651 (560)	-0.45	5300 (550)	-0.30
	CTCAE Fatigue	2711 (582)	-0.33	5782 (568)	-0.20
	CTCAE Flulike syndrome	2711 (582)	-0.33	5782 (564)	-0.20
Social functioning	Performance status	2638 (560)	-0.36	5250 (551)	-0.20
Global quality of life	Performance status	2635 (560)	-0.44	5252 (551)	-0.20
	CTCAE Fatigue	2693 (582)	-0.33	5735 (564)	-0.20
Pain	Performance status	2662 (560)	-0.39	5324 (564)	-0.20
	CTCAE Pain	2724 (582)	-0.61	5824 (571)	-0.32
Fatigue	Performance status	2657 (560)	-0.45	5311 (555)	-0.21
	CTCAE Fatigue	2718 (582)	-0.41	5809 (571)	-0.30
	CTCAE Flulike syndrome	2718 (582)	-0.4	5809 (566)	-0.30
Nausea and/or vomiting	CTCAE Nausea & vomiting	2720 (582)	-0.5	5819 (560)	-0.32
	CTCAE Gastrointestinal	2720 (582)	-0.44	5819 (561)	-0.30
Appetite	Performance status	2648 (560)	-0.43	5291 (560)	-0.20
	CTCAE Anorexia	1086 (226)	-0.49	2440 (222)	-0.31
	Gastrointestinal	2711 (582)	-0.36	5787 (561)	-0.21
Diarrhea	CTCAE Diarrhea	1076 (226)	-0.66	2400 (223)	-0.50
Constipation	CTCAE Constipation	1602 (356)	-0.53	3267 (346)	-0.33

* n_{1R} (n_1) and n_{2R} (n_2) can vary by anchor and EORTC QLQ-C30 scale.

Abbreviations:

EORTC QLQ-C30 = European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire core 30; n_1 = number of patients with at least 1 matched EORTC QLQ-C30 and an anchor form; n_{1R} = number of repeated anchor and HRQOL matched forms across all patients; n_2 = number of patients with at least 2 matched EORTC QLQ-C30 and an anchor form (at least 2 forms are needed to compute change scores); n_{2R} = number of repeated EORTC QLQ-C30 scale and anchor change scores across all patients; CTCAE, common terminology criteria for adverse events.

Table 3: Summary of anchor-based MIDs (weighted average) for both within-group and between-group change over time.

EORTC QLQ-C30 scale	Anchor-based MID for within-group change		Anchor-based MID for between-group difference in change scores	
	MID for Improvement	**MID for Deterioration**	**MID for difference in improvement**	**MID for difference in deterioration**
Physical functioning	6	-8 to -6 (-7)	5	-4
Role functioning	9	-11 to -7 (-9)	7	-8 to -4 (-6)
Social functioning	6	-5	5	-4
Global quality of life	5	-5	4	-4
Pain	8 to 10 (9)	-12	7 to 10 (9)	-9
Fatigue	6	-10 to -8 (-9)	6	-6 to -5 (-6)
Nausea and/or vomiting	no MID	-14 to -13 (-13)	no MID	-9 to -7 (-8)
Appetite loss	6 to 11 (8)	-15 to -8 (-11)	6 to 15 (10)	-9 to -5 (-7)
Constipation	13	-10	13	-9

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

The average MIDs within parenthesis are weighted by the correlations between change score of scale/anchor pairs.

The symptom scores were reversed to follow the functioning scales' interpretation, i.e. 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

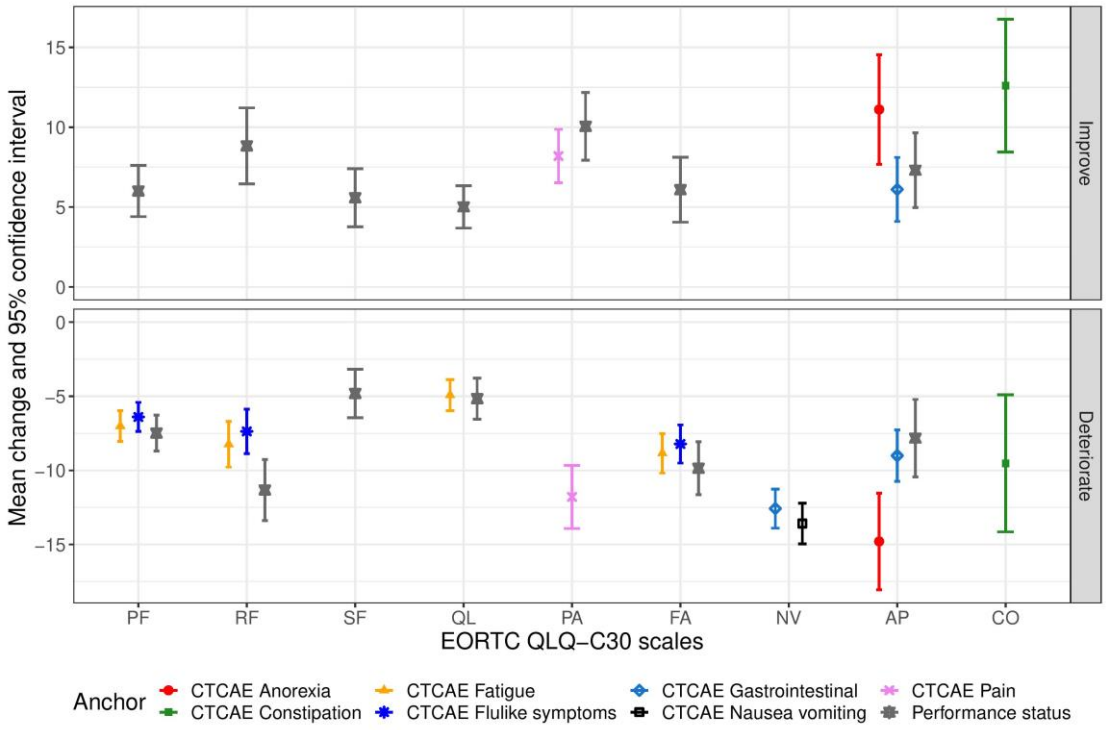

Figure 1: Mean change estimate and 95% confidence interval for improvement and deterioration in EORTC QLQ-C30 scores, across multiple anchors and averaged across different time periods.

Only for scales with at least 1 suitable anchor or with effect size >=0.2 and <0.8 within the "deteriorate" and "improve" groups are plotted.

These mean change estimates are useful for interpreting within-group change in EORTC QLQ-C30 scores over time.

Abbreviations: PF = physical functioning; RF = role functioning; SF = social functioning; QL = global quality of life; PA = pain; FA = fatigue; NV = nausea and/or vomiting; AP = appetite loss; CO = Constipation; 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

1. Appendix

Table A.1: Number of change observations (number of patients) by change scores of suitable anchors

Anchor change score	CTCAE Nausea & vomiting	CTCAE Fatigue	CTCAE Pain	CTCAE Anorexia	CTCAE Flulike symptoms	CTCAE Constipation	CTCAE Gastrointestinal	CTCAE Diarrhoea	Performance status
-4	1 (1)						2 (2)		
-3	30 (10)	16 (5)	14 (7)		14 (5)	15 (4)	20 (8)	1 (1)	
-2	173 (78)	130 (48)	236 (72)	34 (17)	181 (71)	55 (23)	181 (74)	14 (4)	18 (6)
-1	692 (231)	786 (213)	907 (205)	342 (79)	898 (241)	241 (65)	892 (267)	66 (20)	584 (139)
0	3093 (509)	3311 (519)	3969 (531)	1614 (208)	3009 (514)	2812 (334)	2954 (510)	2300 (223)	4012 (524)
1	1304 (336)	1237(321)	580 (178)	386 (98)	1320 (146)	185 (70)	1349 (354)	61 (24)	686 (184)
2	475 (202)	331 (119)	126 (51)	66 (25)	384 (19)	64 (28)	374 (170)	7 (4)	38 (15)
3	68 (39)	29 (16)	8 (5)	13 (3)	34 (5)	5 (4)	65 (37)	6 (2)	3 (2)
4	4 (4)						3 (3)		

Abbreviations: CTCAE, common terminology criteria for adverse events.

Table A.2: Estimated means (effect sizes) of HRQOL change score by clinical change groups that are based on selected anchors per EORTC QLQ-C30 scale

		Mean change method			Linear regression	
EORTC QLQ-C30 scale	**Anchor**	**Improvement (ES)**	**Stable (ES)**	**Deterioration (ES)**	**Improvement**	**Deterioration**
Physical functioning	Performance status	6.05 (0.3)	-2.10 (-0.10)	-7.59 (-0.38)	5.37	-3.70
	CTCAE fatigue	1.09 (0.05) [a]	-2.06 (-0.10)	-7.02 (-0.34)	2.76 [a]	-4.39
	CTCAE flulike symptoms	0.39 (0.02) [a]	-2.21 (-0.11)	-6.41 (-0.31)	2.71 [a]	-3.50
Role functioning	Performance status	8.89 (0.29)	-1.35 (-0.04)	-11.48 (-0.38)	7.36	-7.84
	CTCAE fatigue	0.62 (0.02) [a]	-1.72 (-0.06)	-8.26 (-0.27)	2.08 [a]	-5.75
	CTCAE flulike symptoms	0.86 (0.03) [a]	-2.17 (-0.07)	-7.39 (-0.24)	3.55 [a]	-3.78
Social functioning	Performance status	5.62 (0.22)	-0.79 (-0.03)	-4.87 (-0.20)	4.56	-3.94
Global quality of life	Performance status	5.05 (0.25)	0.2 (0.01)	-5.23 (-0.26)	4.10	-3.85
	CTCAE fatigue	0.99 (0.05) [a]	-0.77 (-0.07)	-4.94 (-0.25)	2.13 [a]	-3.61
Pain	Performance status	10.13 (0.39)	1.9 (-0.07)	-2.67(-0.10) [a]	6.87	-3.54 [a]
	CTCAE pain	8.24 (0.32)	-0.76 (0.)	-11.79 (-0.45)	10.36	-8.99
Fatigue	Performance status	6.13 (0.25)	-2.53(-0.0)	-9.99 (-0.41)	6.22	-5.76
	CTCAE fatigue	3.75 (0.15) [a]	-2.33(-0.0)	-8.88 (-0.36)	5.57 [a]	-6.10
	CTCAE flulike symptoms	3.05 (0.12) [a]	-2.7(-0.0)	-8.24 (-0.34)	5.75 [a]	-4.60
Nausea and/or vomiting	CTCAE gastrointestinal	-0.04 (0.0) [a]	-3.5 (-0.17)	-12.64 (-0.6)	3.86 [a]	-7.18
	CTCAE nausea vomiting	2.43 (0.12) [a]	-3.03 (-0.14)	-13.61 (-0.65)	5.54 [a]	-8.74
Appetite	Performance status	7.36 (0.25)	-2.11 (-0.07)	-7.93 (-0.27)	6.54	-5.43
	CTCAE anorexia	11.11 (0.38)	-5.01 (-0.17)	-14.8 (-0.51)	15.06	-9.09
	CTCAE gastrointestinal	6.13 (0.21)	-1.48 (-0.05)	-9.05 (-0.31)	6.36	-6.21
Diarrhea	CTCAE diarrhea	15.66 (1.34) [a]	0.22 (0.02)	-14.12 (-1.21) [a]	14.36 [a]	-13.25 [a]
Constipation	CTCAE constipation	12.71 (0.57)	-0.74 (-0.03)	-9.52 (-0.43)	12.81	-9.46

[a] These estimated change scores were not considered to summarise the MID estimate because their ES were either <0.2 or ≥0.8
All the ESs for the "stable" group were < 0.2
The symptom scores were reversed to follow the functioning scales interpretation; i.e. 0 represents the worst possible score and 100 the best possible score

Table A.3 Distribution-based estimates

	Distribution-based HRQOL scores at t1 (No. of patients = 602 to 613)			
EORTC QLQ-C30 scale	**0.2 SD**	**0.3 SD**	**0.5 SD**	**1 SEM**
PF	5	7	12	7
RF	7	10	16	14
SF	6	9	15	11
QL	5	7	11	9
FA	5	8	13	11
PA	6	9	16	12
CF	4	6	10	9
EF	5	7	12	9
NV	3	5	8	10
AP	6	10	16	15
DY	6	9	15	13
CO	5	7	12	10
DI	3	5	8	8
SL	7	10	17	15

Abbreviations: t1 is the time point before or on the first day of treatment administration; HRQOL= health-related quality of life; SD = standard deviation; SEM = standard error of measurement; PF = physical functioning; RF = role functioning; CF = cognitive functioning; EF = emotional functioning; SF = social functioning; FA = fatigue; PA = pain; NV = nausea/vomiting; QL = global health status; DY = dyspnoea; AP = appetite loss; SL; sleep disturbance; CO = constipation; DI = diarrhoea