UNIVERSITY OF LEEDS

This is a repository copy of *Minimally important differences for interpreting the EORTC QLQ-C30 in advanced colorectal cancer patients treated with chemotherapy*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/164799/

Version: Accepted Version

Article:
Musoro, JZ, Sodergren, SC, Coens, C et al. (11 more authors) (2020) Minimally important differences for interpreting the EORTC QLQ-C30 in advanced colorectal cancer patients treated with chemotherapy. Colorectal Disease, 22 (12). pp. 2278-2287. ISSN 1462-8910

https://doi.org/10.1111/codi.15295

© 2020 The Association of Coloproctology of Great Britain and Ireland. This is the peer reviewed version of the following article: Musoro, JZ, Sodergren, SC, Coens, C et al. (11 more authors) (2020) Minimally important differences for interpreting the EORTC QLQ-C30 in advanced colorectal cancer patients treated with chemotherapy. Colorectal Disease. ISSN 1462-8910, which has been published in final form at http://doi.org/10.1111/codi.15295. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Minimally important differences for interpreting the EORTC QLQ-C30 in advanced colorectal cancer patients treated with chemotherapy.

Jammbe Z Musoro[1], Sodergren C Samantha[2], Corneel Coens[1], Alessia Pochesci[1], Mitsumi Terada[3], Madeleine T King[4], Mirjam AG Sprangers[5], Mogens Groenvold[6], Kim Cocks[7], Galina Velikova[8], Hans-Henning Flechtner[9] Andrew Bottomley[1] on behalf of the EORTC Quality of Life and Gastro-intestinal Groups

[1]European Organisation for Research and Treatment of Cancer (EORTC), Brussels, Belgium
[2]School of Health Sciences, University of Southampton, Southampton, UK.
[3]Japan Clinical Oncology Group, Clinical Research Support Office and National Cancer Center Hospital, Tokyo, Japan
[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:
Jammbe Musoro, Ph.D., European Organization for Research and Treatment of Cancer, EORTC Headquarters, 83/11 Avenue E. Mounier, 1200 Brussels, Belgium; Tel: +32 (0) 2 774 15 39; jammbe.musoro@eortc.org

ABSTRACT

Background

The European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) assesses health related quality of life of patients in cancer trials. There are currently no minimally important difference (MID) guidelines for the EORTC QLQ-C30 for colorectal cancer (CRC).

Methods

The data were obtained from three published EORTC trials that treated CRC patients using chemotherapy. Potential anchors were selected from clinical variables based on their correlation with EORTC QLQ-C30 scales. Anchor-based MIDs for within-group change and between-group change were estimated via mean change method and linear regression respectively and summarized using weighted correlation. Distribution-based MIDs were also examined.

Results

Anchor-based MIDs were determined for deterioration in 8 of the 14 EORTC QLQ-C30 scales, and in 9 scales for improvement, and varied by scale, direction of change and anchor. MIDs for improvement (deterioration) ranged from 6 to 18 (-11 to -5) points for within-group change and 5 to 15 (-10 to -4) for between-group change. Summarized MIDs (in absolute values) per scale mostly ranged from 5 to 10 points.

Conclusions

These findings have clinical relevance for the interpretation of treatment efficacy and the design of clinical trials by informing sample size requirements.

'What does this paper add to the literature?'

This manuscript determines minimally important differences for interpreting EORTC QLQ-C30 change scores in advanced colorectal cancer. The guidelines will be of clinical interest in terms of assessing treatment efficacy and assisting treatment decision making in colorectal cancer. In addition, the results can inform sample size calculations in future clinical trials.

Keywords: Advanced colorectal cancer, minimally important difference (MID), clinical anchors, health-related quality of life (HRQOL), EORTC QLQ-C30

INTRODUCTION

Globally, colorectal cancer (CRC) is the second and third most commonly diagnosed cancer among women and men respectively [1]. About 20% of patients with CRC are diagnosed in an advanced stage and 50% of newly diagnosed patients will develop advanced disease [2]. Patients with advanced CRC present with locally invasive or metastatic disease or experience recurrence or metastases following treatment. Where surgical resection for advanced CRC is not possible, chemotherapy is recommended to improve symptom control and prolong life. CRC and its treatment lead to specific physical and psychosocial side-effects associated with altered bowel function, stoma placement, and dietary restrictions. The impact of advanced CRC and its treatment on multiple dimensions of life, such as physical, emotional and social, known as health-related quality of life (HRQOL) is recognised as a critical endpoint in cancer clinical trials alongside the traditional measures of treatment response rates and disease-free and overall survival [3, 4]. HRQOL also plays an important role in treatment decision making and features in guidelines such as the ESMO Clinical Practice Guideline for metastatic colorectal cancer (mCRC) [5].

Given the significance of HRQOL, it is imperative that HRQOL data are interpreted in a clinically meaningful manner by establishing what constitutes a minimally important difference (MID) [6, 7, 8, 9, 10, 11]. MID is the smallest difference or change in a HRQOL score that is seen as "important" by a patient or by a third party (e.g. an informed proxy), which might signal a need to modify a patient's management [6]. Common methods for estimating MIDs include anchor-based methods [12, 13, 14] and distribution-based methods [15, 16]. Anchor-based methods rely on external variables that have clinical relevance, such as performance status or toxicity grades or on patient/physician-derived ratings. Distribution-based methods use statistical features, e.g., proportions of the standard deviation (SD) and can be used as supportive evidence to anchor-based methods [11].

The European Organisation for Research and Treatment for Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) [17] is one of the most used instruments in cancer trials [18]. The first recommendations for interpreting EORTC QLQ-C30 in clinical trials were provided by King [7] and Osoba et al. [8]. Based on their overlapping results, mean differences ≥10 points have been commonly considered as clinically meaningful when interpreting EORTC QLQ-C30 scale scores [19]. Nevertheless, there is increasing empirical evidence that MIDs can depend on the HRQOL scale, direction of change scores (improvement versus deterioration) and disease setting [9, 10, 12, 13, 14, 20], implying that a global rule for MIDs applicable to all contexts is highly unlikely [21]. Therefore, there is a need to gather more empirical evidence on MID patterns across scales of the EORTC QLQ-C30 and across disease sites [22]. There are currently no MID guidelines for the EORTC QLQ-C30 specific to CRC. Thus, this study aims to estimate

MIDs for the EORTC QLQ-C30 scales in patients with advanced CRC treated with chemotherapy enrolled in clinical trials. The focus is on establishing MIDs for interpreting HRQOL change scores over time in groups of patients.

METHODS

Description of the data

The study data were obtained retrospectively from three published EORTC trials in advanced CRC. Trial 1 (EORTC 05963) assessed first line infusional 5-fluorouracil, folinic acid and oxaliplatin for metastatic colorectal cancer or loco-regional recurrence and enrolled 564 patients [23]. Trial 2 (EORTC 40952) compared high-dose fluorouracil given as a weekly 24-hour infusion with or without leucovorin versus bolus fluorouracil plus leucovorin in advanced CRC and enrolled 497 patients [24]. Trial 3 (EORTC 40986) compared weekly high-dose infusional fluorouracil plus folinic acid with or without irinotecan in patients with metastatic CRC and enrolled 430 patients [25]. All three trials assessed HRQOL longitudinally using the EORTC QLQ-C30.

The EORTC QLQ-C30

This is a 30-item generic questionnaire that comprises five functioning scales; physical, role, emotional, cognitive and social, three symptoms scales; fatigue, pain, nausea and vomiting and one global health status and quality of life scale. The remaining six single items assess symptoms; dyspnoea, appetite loss, sleep disturbance, constipation, diarrhoea and financial impact. Trial 1 and 2 used version 2 of the EORTC QLQ-C30, whereas trial 3 used version 3. The EORTC QLQ-C30 version 2 and 3 differ only in the response categories of questions 1 to 5, coded as yes/no in version 2, while version 3 uses a four-point Likert scale ranging from 'not at all' to 'very much' for all questions with the exception of the global health status and quality of life which are rated from 1 'very poor' to 7 'excellent'. The scales were scored according to the scoring manual [17], with the means of the raw scores for each scale transformed to fall between 0 and 100. For consistency in signs, all scales were scored such that 0 represents the worst possible score and 100, the best possible score. The financial impact scale was omitted from the analysis.

Clinical anchors

Clinical anchors were selected from available clinical variables such as physician examinations, common terminology criteria for adverse events (CTCAE) and WHO performance status (PS). Anchor selection for each HRQOL scale was based on the correlation strength and clinical

plausibility. Depending on the distribution of the HRQOL scale and anchor pair, a polyserial or polychoric correlation was estimated. Anchors with correlations of ≥|0.30| [11] were given priority. For scales where the majority of the anchors did not reach the 0.3 threshold, we selected anchors with a mixture of weak (<0.3) to optimal correlations. We aimed for multiple anchors per EORTC QLQ-C30 scale to offer some assurance about the plausibility of the MID estimates. All retained anchors were scrutinized for clinical plausibility by five CRC / HRQOL experts to avoid spurious findings. Details on the anchor selection process have been described by Musoro et al. [22].

Definition of clinical change groups

Three clinical change status groups (CCG) were defined: (1) deterioration; worsened by 1 anchor category, (2) stable; no change in anchor category, and (3) improvement; improved by 1 anchor category. Change scores ≥ 2 points in anchor categories were excluded from data sets used to estimate MIDs since they were considered to be clearly above the "minimal" expected change.

Data analysis

Anchor-based methods

Change scores of HRQOL scales and anchors were computed across all pairwise time points and then combined into one dataset to provide adequate data for assessing clinically important changes. As an example, if a patient was measured at time points t_a, t_b and t_c, change scores were computed between t_a & t_b, t_a & t_c and t_b & t_c. This means that a patient can contribute several change scores, and given their change scores, patients can contribute to more than one CCG. Only patients with anchor and HRQOL data for a pair of time points contributed in calculating change scores.

Within-group MIDs, for interpreting change in HRQOL scores over time within a group of patients, were estimated via the mean change method. With this approach, MIDs for improvement and deterioration were calculated as the average HRQOL change scores within the improvement and deterioration CCGs respectively.

Between-group MIDs, for interpreting differences in changes over time between two distinct groups of patients, were estimated using linear regression. For each HRQOL scale and anchor pair, the HRQOL change score was the response variable, and the covariate was a binary anchor variable, 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). Since patients could contribute change scores

to multiple CCGs, and multiple change scores to a specific CCG, we corrected for the association between multiple change scores contributed by some patients via the generalized estimating equations approach [26, 27]. The resulting slope parameters for the 'improved' and 'deteriorated' covariates are the estimated MIDs for improvement and deterioration respectively. Multiple MID estimates per HRQOL scale were triangulated to a single value by using the correlations between change in anchor and HRQOL scores as weights.

To assess whether MIDs varied by age, gender and trial, these factors were included (one at a time), along with their interaction with the binary anchor variable in a regression model. Separate models were fitted for improving and deteriorating HRQOL scores. To account for multiple testing, p-values below 0.001 were considered to be statistically significant.

Distribution-based methods

The 0.2 SD, 0.3 SD and 0.5 SD and standard error of measurement (SEM), were estimated at t1; the time point before or on the first day of treatment administration.

For each CCG, the effect size (ES) was computed as the mean of the HRQOL change scores divided by the SD of the change scores over all time points. According to Cohen's [16] guidelines that an ES of 0.2 is small, 0.5 is moderate and ≥0.8 is large, only mean changes with effect sizes ≥ 0.2 and <0.8 were considered appropriate for inclusion as MID estimates. The rationale here was that an observed ES <0.2 reflects changes that were clinically unimportant, and ESs ≥0.8 were obviously more than minimally important. The SAS software was used for the statistical analysis [28].

RESULTS

In Table 1, a summary of patient and disease characteristics at baseline is presented per trial. A total of 1491 patients were enrolled across the three trials. An overview of patient inclusion in the various analysis phases is presented in Figure A.1.

Twenty-two potential clinical anchors were initially assessed for the EORTC QLQ-C30 scales. After selection of cross-sectional correlations with sufficient magnitude, three to six anchors were identified per HRQOL scale for further review by the clinical panel. The final list of retained anchors (Table 2) comprised PS; scored between 0 (no symptoms of cancer) and 4 (bedbound) and 6 CTCAEs (pain, fatigue, nausea, diarrhoea, flulike syndrome and constipation); with grades ranging from 0 (no toxicity) to 4 (life-threatening). The availability of the anchors varied by trial; CTCAE constipation and fatigue were only available in trial 1, flu-like syndrome in trials 1 and 2, and pain in trials 1 and 3. The remaining anchors; PS CTCAE

diarrhoea and nausea were available in all three trials. At least one clinical anchor was retained for 10 scales of the 14 EORTC QLQ-C30 scales assessed. Table 2 presents the cross-sectional correlations between HRQOL scales and anchors ranging from 0.24 to 0.56 in absolute value, and the correlations between their change scores ranging from 0.13 to 0.33. Table A.1 presents the number of patients and the number of change observations by the anchor change categories. Generally, there were relatively more patients who remained stable than patients who either improved or deteriorated.

Table 3 summarizes anchor-based MIDs (and single value weighted average) for within and between-group change over time for scales for which anchor CCG had an ES of ≥ 0.2 and <0.8. Detailed results are presented in Table A.2. Anchor-based MIDs were determined for deterioration in 8 of the 14 EORTC QLQ-C30 scales assessed, and in 9 scales for improvement. In Figure 1, MIDs from the mean change method from Table 1 are plotted along with their 95% confidence intervals. The figure illustrates how MIDs varied by the HRQOL scale, anchor and, direction of change (improvement versus deterioration). The estimated MIDs followed the expected trends within the various CCGs, i.e. positive change scores within the improvement CCG and negative change scores within the deterioration CCG.

As presented in Table 3, MIDs for assessing change within group ranged from 6.35 to 18.06 points for improvement and -10.66 to -4.83 points for deterioration. Compared to change within groups, MIDs for interpreting change between groups were slightly lower, ranging from 5.43 to 14.56 for improvement and -9.96 to -4.16 deterioration.

Table 3 also presents a correlation-weighted MID average for scales with multiple anchors. The MID averages (including MIDs from scales with single anchors) for deterioration were noticeably smaller in magnitude than improvement for most EORTC QLQ-C30 scales. In general, the correlation-weighted MID averages for most EORTC QLQ-C30 scales were in the range of 6 to 10 points in absolute values for both within and between-group change. The interaction effects between anchor and age, gender and trial respectively showed no statistically significant differences (results not shown).

Compared to the distribution-based estimates, most anchor-based MIDs were > 0.2 SD and were often in the range of 0.3 SD and the 0.5 SD. Distribution-based estimates for all 14 EORTC QLQ-C30 scales considered for this study have been summarised in Table A.3.

DISCUSSION

This study is the first to assess MIDs for interpreting group-level change of EORTC QLQ-C30 scale scores over time in the advanced CRC setting. We determined anchor-based MIDs in 8 of the 14 EORTC QLQ-C30 scales assessed for deterioration, and in 9 scales for improvement. We also provided distribution-based estimates for all 14 EORTC QLQ-C30 scales assessed (Table A.3), which are useful for supporting anchor-based MIDs and aiding interpretation when no anchor-based MIDs are available [11]. We distinguished between anchor-based MIDs for interpreting within-group change over time (e.g. within a treatment group in a trial) obtained from the mean change method, and MIDs for interpreting changes over time between two distinct groups of patients (e.g. treatment versus control group in a trial) obtained from the linear regression.

MIDs are not homogeneous and vary as a result of numerous anchors, different distribution-based measures, and several HRQOL scales. We admit that end users might find such a range of options rather confusing. Hence, we also provide a single MID value per scale (when multiples anchors were used), computed via weighted correlations. We recognize that investigators seeking MIDs may choose to use either the ranges or the single values provided in Table 3.

In general, our MID estimates for most scales, across multiple anchors, were within the range of 5-10 points in absolute values. A similar MID range has been previously reported in patients with breast cancer [8, 20], small-cell lung cancer [8], brain cancer[14], lung cancer[13], and in a pooled data across multiple cancer sites [9,10]. Also in line with previous findings [12, 13, 14, 20], it is important to note that our MIDs varied according to EORTC QLQ-C30 scale and direction of change (improvement versus deterioration). Furthermore, as highlighted by Cocks *et al.* [9, 10], the 5-10 point threshold is not always achievable in all settings. These increasingly robust guidelines reinforce the evidence that scale specific MIDs should be selected with more caution.

Given the increasing interest in using HRQOL scores for managing individual patients, our MIDs can help define cut-offs for individual-level change that are clinically meaningful for CRC patients. For example, in a clinical trial the proportion of patients who change by the MID or more can be computed and compared between treatments. In clinical practice, our MIDs can serve as cut-offs for identifying patients with clinically important problems using HRQOL changes scores. Note that two caveats apply to setting thresholds for use at an individual level. First, the actual threshold needs to be chosen with knowledge of the underlying distribution of each HRQOL scale, since not all MID values will translate into a score that is achievable for an individual. Second, individual thresholds must be set above limits of measurement error to avoid false positive changes that might trigger unnecessary clinical actions [21, 29]. It is important

to indicate that clinical thresholds have also been developed for selected EORTC QLQ-C30 scales to aid individual-level and group-level interpretation in clinical practice [30]. Instead of change scores over time, these thresholds apply to values observed at single visits and were obtained using data from different cancer sites.

In terms of limitations, it remains a challenge to obtain suitable clinical anchors from retrospective clinical trial data. Our anchor-based approach relied on the availability of appropriate anchors in our database. The available clinical anchors were mainly determined by the fact that patients were treated with chemotherapy, and so may not be relevant for patients in surgical colorectal trials. Our data are also limited by the lower prevalence of high grade toxicities, as patient mainly reported grade 0, 1 or 2 toxicities during the trial.

No anchors were found for 4 of the 14 scales assessed; cognitive functioning, social functioning, sleep disturbance and dyspnoea. Furthermore, the available anchors depended solely on clinical interpretations and were mostly not optimally correlated with change scores of HRQOL scales. These low correlations between change scores may be due to measurement error, for instance, Basch et al. [31] reported a lack of concordance between clinician and patient interpretations of what constitutes a change when documenting adverse events.

Our study also lacked anchors that are based on the patient's perspective which, particularly in an advanced setting where treatment is given with palliative intent, the patient's own perception of change is even more vital. Patients' self-assessed rating of change in the HRQOL scores are seldom available in retrospective databases and would need to be embedded within future study protocols to complement the results presented in this paper.

Our data were limited to three published trials, each with particular selection and treatment criteria. The studies considered for this paper primarily included patients with metastatic CRC treated with chemotherapy. Thus, extrapolation outside their specific settings should be done with caution. It is important to highlight that these results form part of a larger project aiming to develop a catalogue of MIDs that is more refined than the single value rule-of-thumb currently still in use. However, we acknowledge that an overly granular approach would be too data-driven and impractical. Thus, in terms of future work, we will further undertake a comprehensive synthesis of MID estimates to identify plausible ranges based on patterns across multiple clinical settings.

In conclusion, our results will support investigators to better understand the value of EORTC QLQ-C30 scores in patients with advanced CRC treated with chemotherapy. Our findings also represent a useful benchmark for judging the success of an intervention, and for calculating sample size in future trials in CRC that use scales of the EORTC QLQ-C30 as endpoints.

Acknowledgements: We thank the members of the EORTC Gastro-intestinal 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.
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.

REFERENCES

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D: Global cancer statistics. CA: a cancer journal for clinicians 2011, 61(2):69-90.
2. Young A, Rea D. ABC of colorectal cancer: treatment of advanced disease. BMJ. 2000;321(7271):1278–1281. doi:10.1136/bmj.321.7271.1278
3. Bottomley A, et al. Health related quality of life outcomes in cancer clinical trials. Eur J Cancer. 2005; 41: 1697-1709.
4. National Institute for Health and Care Excellence. The diagnosis and management of colorectal cancer. Available from: https://www.nice.org.uk/guidance/cg131/evidence/full-guideline-pdf-183509680
5. Van Cutsem E, Cervantes A, Nordlinger B, Arnold D. Metastatic colorectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol (2014) 25 (suppl 3): iii1-iii9.
6. Schünemann HJ, Guyatt GH. Goodbye M(C)ID! Hello MID, where do you come from? Health Serv Res. 2005; 40: 593-597.
7. King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. Qual Life Res. 1996; 5: 555-567.
8. 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.
9. Cocks K, King MT, Velikova G, et al. Evidence-Based Guidelines for Determination of Sample Size and Interpretation of the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. J Clin Oncol 2010; 29(1): 89–96.
10. 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. European Journal of Cancer (2012) 48, 1713– 1721.

11. 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
12. 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
13. 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 Nov; 19(11):1753-60.
14. Maringwa J, et al. Minimal Clinically Meaningful Differences for the EORTC QLQ-C30 and EORTC QLQ-BN20 Scales in Brain Cancer Patients. Ann Oncol. 2011 Sep; 22(9):2107-12.
15. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. J. Clin. Epidemiol. 1999; 52(9), 861–873.
16. Cohen J. Statistical Power Analysis for the Behavioural Sciences (2nd Edition). Lawrence Erlbaum Associates, NJ, USA (1988).
17. 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.
18. Nolte S, Liegl G, Petersen MA, et al. General population normative data for the EORTC QLQ-C30 health-related quality of life questionnaire based on 15,386 persons across 13 European countries, Canada and the Unites States. European Journal of Cancer (2019) 107, 153-163. ISSN 0959-8049https://doi.org/10.1016/j.ejca.2018.11.024
19. 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.
20. Musoro J, Coens C, Fiteni F, et al. EORTC Breast and Quality of Life Groups, Minimally Important Differences for Interpreting EORTC QLQ-C30 Scores in Patients With Advanced Breast Cancer, JNCI Cancer Spectrum, Volume 3, Issue 3, September 2019, pkz037, https://doi.org/10.1093/jncics/pkz037
21. King MT. A point of minimal important difference (MID): a critique of terminology and methods. Expert Rev Pharmacoecon Outcome Res. 2011 Apr; 11(2):171-84.

22. 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
23. Innominato PF, Giacchetti S, Moreau T, et al; ARTBC International Chronotherapy Group. Prediction of survival by neutropenia according to delivery schedule of oxaliplatin-5-Fluorouracil-leucovorin for metastatic colorectal cancer in a randomized international trial (EORTC 05963). Chronobiol Int. 2011 Aug;28(7):586-600. doi: 10.3109/07420528.2011.597532.
24. Köhne C-H., Wils J., Lorenz M. et al. Randomized Phase III Study of High-Dose Fluorouracil Given As a Weekly 24-Hour Infusion With or Without Leucovorin Versus Bolus Fluorouracil Plus Leucovorin in Advanced Colorectal Cancer: European Organization of Research and Treatment of Cancer Gastrointestinal Group Study 40952. Journal of Clinical Oncology 2003 21:20, 3721-3728. DOI: 10.1200/JCO.2003.11.122
25. Köhne C.-H., van Cutsem E., Wils J. et al. Phase III Study of Weekly High-Dose Infusional Fluorouracil Plus Folinic Acid With or Without Irinotecan in Patients With Metastatic Colorectal Cancer: European Organisation for Research and Treatment of Cancer Gastrointestinal Group Study 40986. Journal of Clinical Oncology 2005 23:22, 4856-4865. DOI: 10.1200/JCO.2005.05.546
26. Liang KY, Zeger SL. Regression analysis for correlated data. *Annu. Rev.* Pub Health. 1993;14:43–68.
27. Ying GS, Maguire MG, Glynn R, Rosner B. Tutorial on Biostatistics: Linear Regression Analysis of Continuous Correlated Eye Data. Ophthalmic Epidemiol. 2017;24(2):130-140.
28. Institute Inc. 2013. Base SAS® 9.4 Procedures Guide. Cary, NC: SAS Institute Inc.
29. 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–S45.
30. 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. Published 2016 Jun 7. doi:10.1186/s12955-016-0489-4
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):1043-1050.

Field Code Changed

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

	study			
	EORTC 05963 (N=564)	**EORTC 40952 (N=497)**	**EORTC 40986 (N=430)**	**Total (N=1491)**
	N (%)	**N (%)**	**N (%)**	**N (%)**
Gender				
Male	338 (59.9)	295 (59.4)	268 (62.3)	901 (60.4)
Female	226 (40.1)	196 (39.4)	162 (37.7)	584 (39.2)
Missing	0 (0.0)	6 (1.2)[1]	0 (0.0)	6 (0.4)
Performance status				
0	273 (48.4)	261 (52.5)	246 (57.2)	780 (52.3)
1	231 (41.0)	196 (39.4)	165 (38.4)	592 (39.7)
2	60 (10.6)	33 (6.6)	19 (4.4)	112 (7.5)
Missing	0 (0.0)	7 (1.4)	0 (0.0)	7 (0.5)
Primary tumor site				
Colon	425 (75.4)	260 (52.3)	219 (50.9)	904 (60.6)
Rectum	134 (23.8)	231 (46.5)	210 (48.8)	575 (38.6)
Both (Colon and Rectum)	0 (0.0)	0 (0.0)	1 (0.2)	1 (0.1)
Others	5 (0.9)	0 (0.0)	0 (0.0)	5 (0.3)
Missing	0 (0.0)	6 (1.2)	0 (0.0)	6 (0.4)
Differentiation grade of primary tumor				
Well	0 (0.0)	39 (7.8)	33 (7.7)	72 (4.8)
Moderate	0 (0.0)	331 (66.6)	289 (67.2)	620 (41.6)
Poor	0 (0.0)	101 (20.3)	83 (19.3)	184 (12.3)
Unknown /not available[2]	564 (100.0) [2]	26 (5.2)	25 (5.8)	615 (41.2)
Prior adjuvant chemotherapy				
Yes	101 (17.9)	73 (14.7)	97 (22.6)	271 (18.2)
No	462 (81.9)	418 (84.1)	333 (77.4)	1213 (81.4)
Unknown	1 (0.2)	6 (1.2)	0 (0.0)	7 (0.5)
Prior radiotherapy				
Yes	44 (7.8)	0 (0.0)	30 (7.0)	74 (5.0)
No	517 (91.7)	0 (0.0)	400 (93.0)	917 (61.5)
Unknown /not available[2]	3 (0.5)	497 (100.0) [2]	0 (0.0)	500 (33.5)
Dukes stage				
A	8 (1.4)	0 (0.0)	0 (0.0)	8 (0.5)
B	50 (8.9)	0 (0.0)	0 (0.0)	50 (3.4)
C	85 (15.1)	0 (0.0)	0 (0.0)	85 (5.7)
D	417 (73.9)	0 (0.0)	0 (0.0)	417 (28.0)
Unknown /not available[2]	4 (0.7)	497 (100.0) [2]	430 (100.0) [2]	931 (62.4)
M classification				
1	0 (0.0)	278 (55.9)	430 (100.0)	708 (47.5)
0	0 (0.0)	149 (30.0)	0 (0.0)	149 (10.0)
Unknown /not available[2]	564 (100.0) [2]	70 (14.1)	0 (0.0)	634 (42.5)
Country				
Germany	2 (0.4)	398 (80.1)	290 (67.4)	690 (46.3)
Belgium	100 (17.7)	6 (1.2)	66 (15.3)	172 (11.5)
France	149 (26.4)	0 (0.0)	23 (5.3)	172 (11.5)

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

	study			
	EORTC 05963 (N=564)	**EORTC 40952 (N=497)**	**EORTC 40986 (N=430)**	**Total (N=1491)**
	N (%)	**N (%)**	**N (%)**	**N (%)**
Italy	156 (27.7)	4 (0.8)	0 (0.0)	160 (10.7)
Canada	80 (14.2)	0 (0.0)	0 (0.0)	80 (5.4)
Netherlands	0 (0.0)	52 (10.5)	28 (6.5)	80 (5.4)
Egypt	0 (0.0)	13 (2.6)	21 (4.9)	34 (2.3)
Norway	28 (5.0)	0 (0.0)	0 (0.0)	28 (1.9)
Others	49 (8.7)	13 (2.6)	2 (0.5)	64 (4.3)
Age				
Mean (SD)	59.86 (10.24)	59.69 (10.10)	60.25 (9.84)	59.92 (10.07)
Interquartile	53.0 - 68.0	55.0 - 67.0	54.0 - 68.0	54.0 - 68.0

[1]Ineligible or lost to follow-up

[2]Other countries: Austria, Portugal, Russia, United Kingdom, Greece, South Africa and Israel

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

		Score		Change score	
Scale	**Anchor**	**n_1 (n_{1R})***	**Correlation**	**n_2 (n_{2R})***	**Correlation**
PF	Performance status	1210 (3816)	-0.40	763 (6433)	-0.30
	CTCAE Fatigue	439 (1100)	-0.30	280 (1001)	-0.21
RF	Performance status	1202 (3766)	-0.40	763 (6157)	-0.30
	CTCAE Fatigue	438 (1087)	-0.40	280 (948)	-0.30
SF	Performance status	1210 (3792)	-0.33	763 (6242)	-0.23
	CTCAE Fatigue	438 (1099)	-0.40	280 (1001)	-0.25
QL	Performance status	1202 (3772)	-0.34	740 (6220)	-0.23
	CTCAE Flulike syndrome	777 (1935)	-0.24	500 (2658)	-0.20
	CTCAE Fatigue	436 (1089)	-0.31	280 (980)	-0.20
PA	CTCAE Pain	796 (2152)	-0.45	561 (3288)	-0.30
FA	Performance status	1210 (3807)	-0.41	763 (6289)	-0.30
	CTCAE Fatigue	440 (1103)	-0.40	280 (1005)	-0.31
NV	CTCAE Nausea	1139 (3003)	-0.41	754 (4632)	-0.20
	CTCAE Flulike syndrome	782 (1960)	-0.24	500 (2712)	-0.20
AP	Performance status	1210 (3797)	-0.43	763 (6242)	-0.25
DI	CTCAE Diarrhea	1135 (2969)	-0.42	754 (4516)	-0.13
CO	CTCAE Constipation	1139 (2979)	-0.56	307 (1344)	-0.33

* n1 (n1R) and n2 (n2R) 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; n1 = number of patients with at least 1 matched EORTC QLQ-C30 and an anchor form; n1R = number of repeated anchor and HRQOL matched forms across all patients; n2 = 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); n2R = number of repeated EORTC

QLQ-C30 scale and anchor change scores across all patients; 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; DI= diarrhoea; CO = Constipation; CTCAE, common terminology criteria for adverse events.

422
423

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

	Anchor-based MID for within-group change		Anchor-based MID for between-group difference in change	
Scale	**Improvement**	**Deterioration**	**Improvement**	**Deterioration**
PF	7.31 to 8.52 (7.81)	-8.43 to -6.09 (-7.47)	6.05 to 10.04 (7. 69)	-7.23 to -4.16 (-5.96)
RF	10.43 to 18.06 (14.24)	-10.66	7.95 to 14.17 (11.06)	-9.96
SF	8.11 to 10.26 (9.23)	-6.18	6.73 to 7.79 (7.28)	-6.03
QL	7.14 to 10.34 (8.43)	-7.97 to -4.83 (-6.38)	5.53 to 6.36 (5.86)	-9.12 to -6.81 (-8.13)
FA	7.65 to 13.82 (10.79)	-7.73 to -7.05 (-7.38)	5.43 to 12.01 (8.77)	-6. 98 to -6.76 (-6.87)
NV	7.75	-7.95 to 5.30 (-6.62)	7.34	-7.33 to -5.17 (-6.25)
AP	12.28	-9.78	10.0	-7.11
DI	6.35	-7.96	8.25	-5.46
CO	12.75	No MID	14.56	No MID

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

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; DI= diarrhoea; CO = Constipation; CTCAE, common terminology criteria for adverse events.

424 **Figure 1**

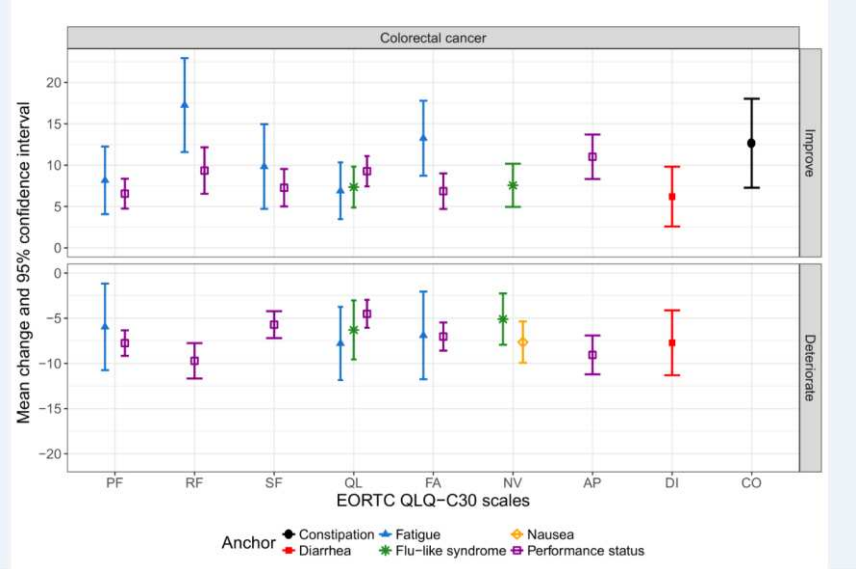
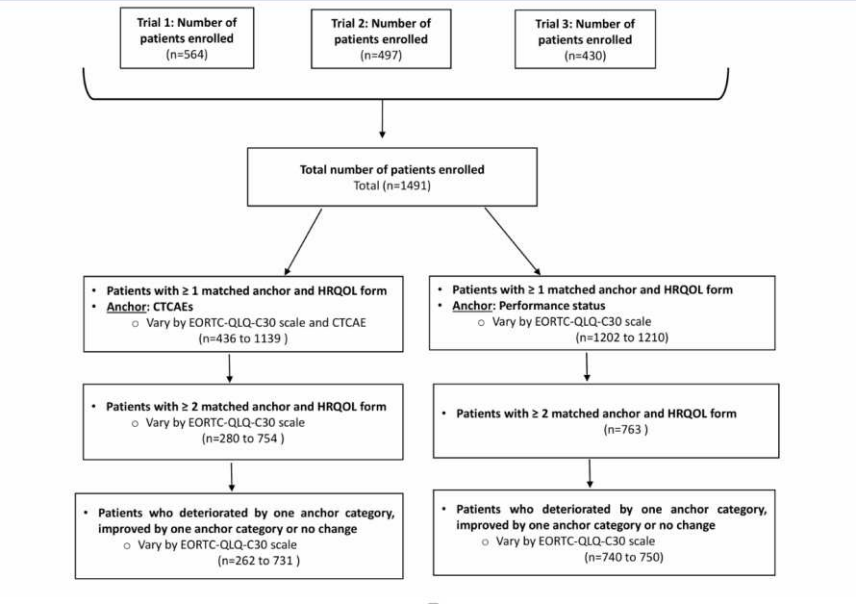

Colorectal cancer
Improve
Deteriorate
Mean change and 95% confidence interval
20
15
10
5
0
0
−5
−10
−15
−20
PF
RF
SF
QL
FA
NV
AP
DI
CO
EORTC QLQ−C30 scales
Anchor
Constipation
Fatigue
Nausea
Diarrhea
Flu−like syndrome
Performance status

Figure 2


Trial 1: Number of patients enrolled (n=564)
Trial 2: Number of patients enrolled (n=497)
Trial 3: Number of patients enrolled (n=430)
Total number of patients enrolled
Total (n=1491)
Patients with ≥ 1 matched anchor and HRQOL form
Anchor: CTCAEs
Vary by EORTC-QLQ-C30 scale and CTCAE
(n=436 to 1139)
Patients with ≥ 1 matched anchor and HRQOL form
Anchor: Performance status
Vary by EORTC-QLQ-C30 scale
(n=1202 to 1210)
Patients with ≥ 2 matched anchor and HRQOL form
Vary by EORTC-QLQ-C30 scale
(n=280 to 754)
Patients with ≥ 2 matched anchor and HRQOL form
(n=763)
Patients who deteriorated by one anchor category, improved by one anchor category or no change
Vary by EORTC-QLQ-C30 scale
(n=262 to 731)
Patients who deteriorated by one anchor category, improved by one anchor category or no change
Vary by EORTC-QLQ-C30 scale
(n=740 to 750)