



The european organisation for research and treatment of cancer head and neck cancer module (EORTC QLQ-HN43): Estimates for minimal important difference and minimal important change

Susanne Singer^{a,*}, Eva Hammerlid^{b,2}, Iwona M. Tomaszewska^{c,3}, Cecilie D. Amdal^{d,e,4}, Bente B. Herlofson^{f,5}, Marcos Santos^{g,6}, Joaquim Castro Silva^{h,7}, Hisham Mehanna^{i,8}, Amy Fullerton^{j,9}, Teresa Young^{k,10}, Loreto Fernandez Gonzalez^{l,11}, Johanna Inhestern^{m,12}, Monica Pinto^{n,13}, Juan I. Arraras^{o,14}, Noam Yarom^{p,q,15}, Pierluigi Bonomo^{r,16}, Ingo Baumann^s, Razvan Galalae^{t,17}, Ourania Nicolatou-Galitis^{u,18}, Naomi Kiyota^{v,19}, Judith Raber-Durlacher^{w,20}, Dina Salem^{x,21}, Alexander Fabian^{y,22}, Andreas Boehm^{z,23}, Sanja Krejovic-Trivic^{aa,24}, Wei-Chu Chie^{ab,25}, Katherine J. Taylor^{a,26}, Allen C. Sherman^{ac,27}, Lisa Licitra^{ad,28}, Jean-Pascal Machiels^{ae,29}, Kristin Bjordal^{e,af,30}, on behalf of the EORTC Quality of Life Group and the EORTC Head and Neck Cancer Group

^a Division of Epidemiology and Health Services Research, Institute of Medical Biostatistics, Epidemiology, and Informatics (IMBEI), University Medical Centre of Johannes Gutenberg University, Mainz, Germany

^b Department of Otorhinolaryngology-Head and Neck Surgery, Institute of Clinical Sciences, Sahlgrenska Academy at University of Gothenburg, Sahlgrenska University Hospital, Gothenburg, Sweden

^c Department of Medical Didactics, Jagiellonian University Medical College, Krakow, Poland

^d Department of Oncology, Oslo University Hospital, Norway

^e Department of Research Support Services, Oslo University Hospital, Norway

^f Department of Oral Surgery and Oral Medicine, University of Oslo, and Department of Otorhinolaryngology - Head and Neck Surgery Division for Head, Neck and Reconstructive Surgery, Oslo University Hospital, Oslo, Norway

^g Radiation Oncology Department, Grupo CONFIAR, Goiania, GO, Brazil

^h Department of Otorhinolaryngology, Head and Neck Surgery, Instituto Português de Oncologia Francisco Gentil do Porto, Porto, Portugal

ⁱ Institute of Head and Neck Studies and Education, University of Birmingham, Birmingham, UK

^j Department of Communication Sciences and Disorders, Brooks Rehabilitation College of Healthcare Sciences, Jacksonville University, Jacksonville, FL, USA

^k Lynda Jackson Macmillan Centre, East & North Hertfordshire NHS Trust incorporating Mount Vernon Cancer Centre, Northwood, UK

^l Instituto Oncologico Fundación Arturo Lopez Perez, Santiago, Chile

^m Department of Otorhinolaryngology, Oberhavelkliniken, Hennigsdorf, Germany

ⁿ Strategic Health Services Department, Istituto Nazionale Tumori -IRCCS, Fondazione G. Pascale, Napoli, Italy

^o Oncology Departments, Complejo Hospitalario de Navarra, Pamplona, Spain

^p Oral Medicine Unit, Sheba Medical Center, Tel-Hashomer, Israel

^q The Maurice and Gabriela Goldschleger School of Dental Medicine, Faculty of Medical and Health Sciences, Tel Aviv University, Tel Aviv, Israel

^r Radiation Oncology, Azienda Ospedaliero-Universitaria Careggi, Florence, Italy

^s Department of Otorhinolaryngology, Head and Neck Surgery, University of Heidelberg, Heidelberg, Germany

^t Heavy Ion Radiotherapy Center - MedAustron, Vienna, Austria

^u Clinic of Hospital Dentistry, Dental Oncology Unit, School of Dentistry, National and Kapodistrian University of Athens, Athens, Greece

^v Department of Medical Oncology and Hematology, Kobe University Hospital Cancer Center, Kobe, Japan

^w Department of Oral and Maxillofacial Surgery, Amsterdam University Medical Center, University of Amsterdam, and Department of Oral Medicine ACTA, University of Amsterdam and Vrije Universiteit, Amsterdam, the Netherlands

^x Department of Medical Oncology, Ain Shams-University, Cairo, Egypt

^y Department of Radiation Oncology, University Hospital Schleswig-Holstein, Kiel, Germany

^z Department of Otorhinolaryngology Head and Neck Surgery, St. Georg Hospital, Leipzig, Germany

^{aa} Clinic of Otorhinolaryngology and Maxillofacial Surgery, Faculty of Medicine, University of Belgrade, Belgrade, Serbia

^{ab} Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taiwan

^{ac} Behavioral Medicine Division, Winthrop P. Rockefeller Cancer Institute, University of Arkansas for Medical Sciences, Little Rock, USA

^{ad} Department of Medical Oncology, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy

* Correspondence to: University Medical Centre of Johannes Gutenberg University, Institute of Medical Biostatistics, Epidemiology, and Informatics, Division of Epidemiology and Health Services Research, Obere Zahlbacher Straße 69, 55131 Mainz, Germany.

E-mail address: susanne.singer@uni-mainz.de (S. Singer).

<https://doi.org/10.1016/j.ejca.2024.115062>

Received 23 August 2024; Received in revised form 2 October 2024; Accepted 3 October 2024

Available online 9 October 2024

0959-8049/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

ARTICLE INFO

Keywords:

Minimal important change
Minimal important difference
Minimal detectable difference
Clinical significance
Subjective significance
EORTC QLQ-C30
EORTC QLQ-HN43, quality of life
Patient-reported outcomes

ABSTRACT

Introduction: Minimal important change estimates (MIC) are useful for interpreting results of clinical research with quality of life (QoL) as an endpoint. For the European Organisation for Research and Treatment of Cancer head and neck cancer module, the EORTC QLQ-HN43, no such thresholds are established.

Methods: Head and neck cancer patients under active treatment ($n = 503$) from 15 countries completed the EORTC QLQ-HN43 three times (t1: before treatment, t2: three months after t1, t3: six months after t1). A subgroup completed a Subjective Significance Questionnaire (SSQ), indicating experienced change from the previous time point in four QoL domains. QoL was assumed to deteriorate after t1 and improve again until t3. The MIC was established using the average of mean differences in SSQ groups (MIC_{mean}) and estimates based on logistic regressions ($MIC_{predict}$). Additionally, minimal detectable changes (MDC) were computed using 0.5 standard deviation and standard error of the mean.

Results: For *swallowing*, *speech*, *dry mouth*, and *global QoL*, the MIC for deterioration were 13, 14, 26, and 10 respectively. The MIC for improvement were 8 (*swallowing*), 6 (*dry mouth*), and 5 (*global QoL*); no MIC for *speech* improvement can be presented because of insufficient correlation between change score and anchor. The MDC estimates for deterioration were 15, 14, 15, and 11. For improvement, the MDC estimates were 13, 14, 14, and 11.

Conclusions: Our results underline that no single MIC or MDC can be applied to all EORTC QLQ-HN43 scales, and that the MIC for deterioration seems larger than those for improvement.

1. Introduction

One method to estimate the effect of a treatment is to compute the percentage of responders. To do this, however, what constitutes a responder needs to be defined. For patient-reported outcomes, a common method to define a “response” in the score of a specific questionnaire is to calculate its Minimal Important Difference (MID) or Minimal Important Change (MIC).

The MID is the minimal difference *between groups* that is considered to be of clinical relevance. It is often defined as the smallest score difference that patients consider meaningful or that would prompt a discussion with the patient in relation to the standard treatment or

supportive care options [1–3]. There is no established gold standard method for calculating the MID, but it is common to use both anchor-based and distribution-based approaches (though the latter one should rather be named Minimal Detectable Difference, MDD, because it is conceptually different to the MID [1,4,5], which is why we use this term).

The MIC has been defined as the “threshold for a minimal within-person change over time above which patients perceive themselves importantly changed” [5]. This definition precludes using distribution-based methods for its calculation because measures identified by this approach should rather be called Minimal Detectable Change (MDC). It also excludes using third parties (physicians, parents etc.) to define what change is important. Other authors [2,6] use the term MIC more broadly for all least *within-person changes* that are of clinical relevance, thereby including both anchor- and distribution-based approaches and allowing third parties to define the anchor. In our study, we use the term “minimal important” (MIC/MID) for anchor-based approaches and “minimal detectable” (MDC/MDD) for distribution-based approaches [7].

There are different opinions in the scientific community about the equivalence of distribution-based vs. anchor-based approaches. The latter ones are often preferred, but this is not always the case. Crosby [2] argued that a combination and integration of both approaches should be considered. We decided to use both approaches but consider anchor-based estimates (MIC/MID) to be clinically more relevant.

One of the most widely used questionnaires to measure health-related quality of life (HRQoL) in patients with cancer is the European Organisation for Research and Treatment of Cancer Quality of Life Core Questionnaire (EORTC QLQ-C30) [8] and its accompanying modules for specific cancer entities or conditions, such as for head and neck cancer [9,10] or fatigue [11].

For the EORTC QLQ-C30, a number of studies have been conducted to establish its MID [12–19] and MIC [20–22]. However, for its modules, this has rarely been done. Sully et al. investigated it for the myeloma module [23], Reni et al. for the pancreatic cancer module [24], and Coomans et al. for the brain cancer module [25].

We recently updated the head and neck cancer module, the EORTC QLQ-HN43 [10,26,27], and explored various approaches to estimate its MID and MIC [28], using its *swallowing* scale as an example. This exploration was necessary because there are many methods available to define MID and MIC, without clear advantages of one over another.

¹ ORCID 0000-0001-5784-7964² ORCID 0000-0002-1813-108X³ ORCID 0000-0002-4118-5596⁴ ORCID 0000-0001-8422-8085⁵ ORCID 0000-0002-9621-9263⁶ ORCID 0000-0002-7055-4397⁷ ORCID 0000-0002-3934-9950⁸ ORCID 0000-0002-5544-6224⁹ ORCID 0000-0002-4336-5523¹⁰ ORCID 0000-0003-0367-2980¹¹ ORCID 0000-0001-5026-6438¹² ORCID 0000-0003-4613-2810¹³ ORCID 0000-0002-0823-4907¹⁴ ORCID 0000-0002-4415-2570¹⁵ ORCID 0000-0001-9771-7857¹⁶ ORCID 0000-0001-8999-5208¹⁷ ORCID 0000-0001-8800-5460¹⁸ ORCID 0000-0001-9036-6329¹⁹ ORCID 0000-0001-8021-6116²⁰ ORCID 0000-0003-1500-8390²¹ ORCID 0000-0001-9281-0999²² ORCID 0000-0003-2415-527X²³ ORCID 0000-0001-9511-9167²⁴ ORCID 0000-0002-2160-6357²⁵ ORCID 0000-0001-5584-6554²⁶ ORCID 0000-0003-0768-5789²⁷ ORCID 0000-0001-5457-1659²⁸ ORCID 0000-0003-0623-4118²⁹ ORCID 0000-0001-6369-9742³⁰ ORCID 0000-0002-1914-494X

While the mean difference in score between two time points among patients who say their HRQoL has changed “a little” is a method often used, there are also suggestions to use receiver operating curves (ROC) [29] or logistic regression [5,6,30,31]. We found that the ROC approach did not work well in our sample, whereas subjective change ratings and predictive modelling for establishing the MIC and standard deviation based approaches for establishing the MDD/MDC were suitable methods [28].

In this paper, we present the MID/MDD and MIC/MDC estimates using these methods for all scales of the EORTC QLQ-HN43, together with the ones of the EORTC QLQ-C30, for patients with head and neck cancer. This study represents the first effort to examine these important thresholds for a widely used measure, using a large international sample of patients.

2. Methods

2.1. Study design and data collection

This was a multi-national prospective cohort study with consecutive accrual. Its aim was to determine key psychometric properties of the EORTC QLQ-HN43. Patients included were about to start treatment, 18 years or older, able to complete a questionnaire and had one of the following diagnoses: lip, oral cavity, salivary glands, oro-hypopharynx, nasopharynx, nasal cavity, nasal sinuses, larynx, sarcoma in the head and neck region, or lymph node metastases from unknown primary in the head and neck area.

Patients completed the questionnaire before treatment (“baseline” - t1), three months after baseline (t2), and six months after baseline (t3). It was assumed that HRQoL would on average deteriorate after t1 and improve again until t3, based on findings in the literature [32] and on expertise from the clinicians in our group.

Collaborators in the following countries enrolled patients: Brazil, Chile, Egypt, Germany, Greece, Israel, Italy, Japan, Norway, Poland, Portugal, Spain, Sweden, the United Kingdom, and the United States.

2.2. Instruments

The questionnaires for which we sought to determine the MIC and MID were the EORTC QLQ-C30 [8] and its head and neck cancer module, the EORTC QLQ-HN43 [10]. The EORTC QLQ-C30 consists of 30 items, combined into five functional and nine symptom scales. Recently, it has been suggested that a summary score can also be meaningfully used [33,34]. The 43 items of the EORTC QLQ-HN43 can be combined into 19 scales [10].

At time points t2 and t3, in some centres, participants were also asked to complete the Subjective Significance Questionnaire (SSQ) [35]. In the SSQ, the patients rated the extent to which their HRQoL had changed (improved or worsened) in the areas of four selected domains: *swallowing*, *speech*, *dry mouth* and *overall*³¹ *quality of life* compared to the previous time point (e.g., “Since the last time I filled out the questionnaire, my swallowing is ...”). The response options for these items ranged from very much worse to very much better on a 7-point Likert scale. We did not ask the participants to judge the change of their HRQoL in all domains in order to keep the questionnaire burden low. The first three domains were selected as they have previously been rated as most important by patients with head and neck cancer [27], and overall quality of life was included due to its general applicability. The SSQ served as anchor for important change.

The collaborators documented information on gender, age, education, tumour location, tumour stage at baseline. The treatment and

Karnofsky Performance Score (KPS) were documented at each time point.

2.3. Analysis

We developed the analysis plan through iterative discussions and based on the results of our previous work for the swallowing scale [28]. Guidance from Canadian [36], US American [2], and Dutch [4–6,37] working groups were also considered.

2.3.1. Descriptive analyses

The mean change scores (delta Δ) between t1 and t2 as well as between t2 and t3 were calculated for all scales together with their standard deviation (SD) and minimum/maximum. The proportion of patients improved and deteriorated based on the anchor was obtained for each SSQ item.

2.3.2. Methods for estimating the minimal important difference (MID)

We calculated Pearson correlation coefficients between the KPS and EORTC scales at each time point. It was defined a priori that if a correlation was ≥ 0.70 , the KPS would be used as an anchor for the MID, keeping in mind that this would be using a relatively rough physician rating to define meaningful differences for patient-reported outcome measures, which is conceptually weak. However, due to the lack of other suitable anchors available in the database we decided to try this approach but use a relatively high threshold [38]. Correlations below 0.70 were considered to indicate that the KPS was not an adequate anchor. According to Devji et al. [36], a correlation of ≥ 0.70 indicates “definitely” good correlation of anchor and PROM, supporting the credibility of the MID estimate.

2.3.3. Methods for estimating the minimal detectable difference (MDD)

For each EORTC scale, we computed the 0.5 SD and the standard error of measurement (SEM) of the scale score at t1, t2 and t3. The SEM was obtained by calculating $SEM = SD \times \text{square root of } (1 - \text{Cronbach's } \alpha)$. The alpha was derived from the current sample, specifically for each time point. We then averaged the estimates obtained from all time points and both approaches (SEM and SD) for each EORTC scale.

2.3.4. Methods for estimating the minimal important change (MIC)

We used the following steps to estimate the MICs:

11. MIC for deterioration:

- a. The mean EORTC change score of the subset of patients who reported that they had deteriorated “a little” between t1 and t2 on the relevant SSQ item was calculated (the MIC_{mean}). That is, we examined the mean change in EORTC *swallowing*, *speech*, *dry mouth*, and *global health status/QoL* scores for the subgroup of patients who reported “a little” deterioration on the t2 SSQ *swallowing*, *speech*, *dry mouth*, and *overall quality of life* items, respectively. In addition, as a secondary analysis, we examined the mean change in each of the other EORTC scales for patients who reported that they were “a little worse” on the SSQ *overall quality of life* item at t2. These secondary analyses examined the viability of an indirect general anchor (subjective changes in *overall quality of life*) to evaluate changes in specific domains (*senses*, *shoulder problems*, *coughing*, etc.) by point-biserial correlations of the anchor with the EORTC change score. We anticipated that findings would be less rigorous or precise. This method of estimating the MIC by using mean change scores for the subgroup of patients who endorse a category that represents minimal change on an anchor (e.g., “a little worse” on the SSQ) has several drawbacks (imprecise estimates due to small samples and systematic bias towards zero [5]) but this is a very commonly used approach and it is displayed for comparison with other studies.

³¹ The respective scale of the EORTC QLQ-C30 is named “global health status/quality of life” but for the SSQ, we used the original wording developed by Osoba et al. [33] which is “overall quality of my life”.

- b. The responses per SSQ item measured at t2 were grouped into a *deteriorated* category (including the responses “very much worse”, “moderately worse”, “a little worse”) versus a *not deteriorated* category (including “about the same”, “a little better”, “moderately better”, “very much better”). Using these two categories, we calculated a logistic regression predicting SSQ “deterioration” (vs. “not deteriorated”) between t1 and t2, with the EORTC change score between t1 and t2 as the independent variable, and obtained the MIC based on predictive modelling (the MIC_{predict}) by calculating $(\ln(\text{odds-pre-intercept})/\text{regression coefficient})$ [6]. Separate logistic regression models were tested predicting “deterioration” vs “no deterioration” at t2 on the SSQ *swallowing*, *speech*, *dry mouth*, and *overall quality of life* items from change scores on the EORTC *swallowing*, *speech*, *dry mouth*, and *global health status/QoL* scales, respectively. In addition to these analyses that used a specific conceptually relevant anchor for each EORTC change score, we also conducted secondary analyses that examined the viability of an indirect, general anchor (the SSQ *overall quality of life* item) for each of the other EORTC scales. Again, we expected the results of these secondary analyses to be of interest but less compelling.

12. MIC for improvement:

- a. The mean EORTC change score of the subset of patients who reported that they had improved “a little” between t2 and t3 on the relevant SSQ item was calculated. That is, we examined the mean change in EORTC *swallowing*, *speech*, *dry mouth*, and *global health status/QoL* scales for the subgroup of patients who reported “a little” improvement on the t3 SSQ *swallowing*, *speech*, *dry mouth*, and *overall quality of life* items, respectively. In ancillary analyses, we used the indirect global anchor, evaluating mean change in each of the other EORTC scales for the subgroup of patients who reported “a little” improvement in the SSQ overall QoL item at t3.
- b. SSQ responses from t3 were grouped into an *improved* category (“very much improved” to “a little improved”) versus a *not improved* category (“about the same” and all indications of worsening). Predictive modelling was performed analogous to the procedure for deterioration, but this time with improvement at t3 as the outcome variable and the changes between t2 and t3 taken as covariate. Separate logistic regression analyses were used to predict “improvement” vs “no improvement” at t3 on the SSQ *swallowing*, *speech*, *dry mouth*, and *global quality of life* items from change scores on the EORTC *swallowing*, *speech*, *dry mouth*, and *global health status/QoL* scales, respectively. In addition to the analyses that used a specific conceptually relevant anchor for each EORTC score, we also conducted secondary analyses that examined the viability of the general anchor—the SSQ overall QoL item at t3, for each of the other EORTC scales. If the percentage of improved patients was < 0.40 for a SSQ anchor, the MIC_{predict} was corrected with the formula $\text{MIC}_{\text{ad-justed}} = \text{MIC}_{\text{predict}} - (0.09 + 0.103 \times \text{point biserial correlation coefficient of change score with anchor}) \times \text{SD of change score} \times \text{log-odds (imp)}$ [5,37].

To calculate the MIC for each EORTC scale, we averaged together the estimates obtained from the MIC_{mean} and MIC_{predict} approaches, except for those scales for which the association of the anchor with the change score was inadequate, according to guidance provided by Terwee, although with a lower threshold [5]. This was assessed by calculating point biserial correlations between the anchor (the dichotomized SSQ items) and the relevant EORTC change scores. Correlation coefficients below 0.15 were considered to be inadequate and thus no average was calculated [5]. Notably, this threshold was not used to decide whether the MIC should be calculated or not, but rather whether the two MIC estimates (MIC_{mean} and MIC_{predict}) were combined or not. All MIC and the corresponding correlation coefficients are presented. Readers can

decide for their specific research purpose whether the correlation seems to be sufficient or not.

2.3.5. Methods for estimating the minimal detectable change (MDC)

For each EORTC scale, we calculated the SEM and the 0.5 SD of the Δ between t1 and t2 (when most QoL scores were expected to become worse) to determine the MDC for *deterioration* as well as the SEM and the 0.5 SD of the Δ between t2 and t3 (when most scores were expected to show some recovery) to determine MDC for *improvement*.

The SEM for the MDC was obtained by calculating $\text{SEM} = \text{SD} \times \text{square root of } (1 - \text{intra-class correlation coefficient (ICC) between t1 and t2})$. Here, the ICC rather than Cronbach’s alpha was used as the reliability coefficient because the MDC is a measure of change and a test-retest-measure is more appropriate [4]. The ICC was derived from our sample.

3. Results

3.1. Participants

There were 503 participants from 15 countries (Table 1). They were 19 to 87 years old (mean 63, SD 11 years). The patient flow is presented in Figure 1 and is described in detail elsewhere [28]. The SSQ was completed by 213 patients at t2 and by 214 patients at t3.

3.2. Changes in quality of life and proportion improved / deteriorated

The mean EORTC scale scores together with SD are displayed at eTable 1 (supplement). Average EORTC change scores (Δ) between t1

Table 1
Patient characteristics (n = 503).

Variable	Category	N	Percent
Gender	Male	365	73 %
	Female	134	27 %
	Missing information	4	1 %
Education	< 10 years	157	31 %
	10 years	83	17 %
	> 10 years	223	44 %
	unknown	40	8 %
Cancer site	Larynx	81	16 %
	Hypopharynx	44	9 %
	Oropharynx	173	34 %
	Oral cavity	151	30 %
	Parotid gland	22	4 %
	Nasal cavity and sinuses	17	3 %
	Unknown primary	13	3 %
	Missing information	2	0.4 %
Cancer stage	I	59	12 %
	II	90	18 %
	III	93	18 %
	IV	249	50 %
	Missing information	12	2 %
Karnofsky Performance Score(at baseline)	100	180	36 %
	90	174	35 %
	80	87	17 %
	70	40	8 %
	60	11	2 %
	50	1	< 1 %
	40	2	< 1 %
	Missing information	8	2 %
Treatment	Surgery alone	95	19 %
	Radiotherapy alone	86	17 %
	Chemotherapy alone	11	2 %
	Chemo-radiation	136	27 %
	Chemo-radiation with surgery	74	15 %
	Surgery and chemotherapy only	6	1 %
	Surgery and radiotherapy only	68	14 %
	Other	27	5 %

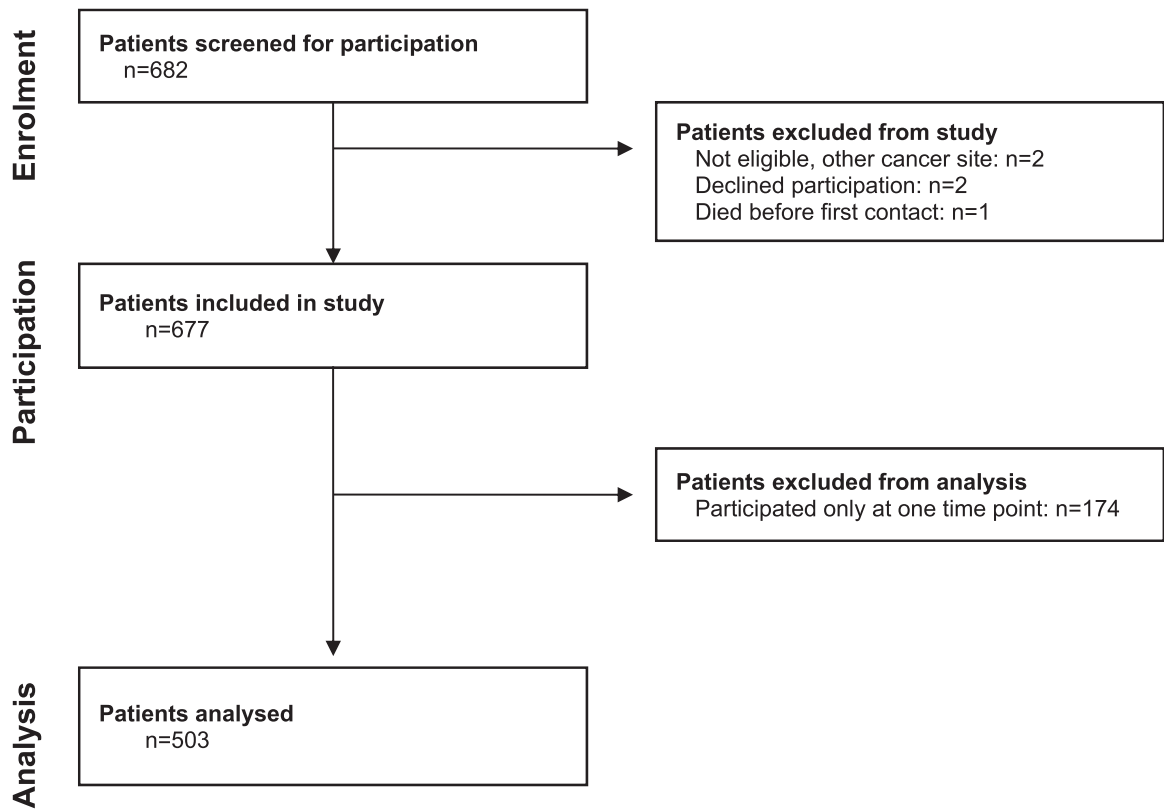


Fig. 1. Patient flow through the study.

and t2 and between t2 and t3 together with 95 % confidence intervals can be found in eTable 2.

The proportion of patients who reported any level of deterioration (i.e., very much worse, moderately worse, or a little worse) between t1 and t2 for the four SSQ items ranged from 42 % (speech) to 64 % (dry mouth). The proportion of patients who experienced any level of improvement (i.e., very much better, moderately better, or a little better) between t2 and t3 ranged from 31 % (dry mouth) to 51 % (overall quality of life) (Table 2).

3.3. Minimal important difference (MID)

The correlation coefficients of the various EORTC scale scores with the KPS turned out to be below our threshold of 0.70 (eTable 3) and,

therefore, KPS was disregarded as an anchor.

3.4. Minimal detectable difference (MDD)

The MDD, taken as the average of SEM and SD-based estimates of all time points combined, ranged from 9, for nausea and vomiting, to 21, for insomnia, appetite loss, and role functioning (Table 3).

3.5. Minimal important change (MIC)

The correlation coefficients between the four specific SSQ anchors (i.e., swallowing, speech, dry mouth, overall quality of life) and the respective EORTC change scores ranged from 0.30 to 0.49 regarding deterioration, and from 0.15 to 0.42 regarding improvement (Table 4).

The correlation with the SSQ global anchor “overall quality of life” was ≥ 0.30 for three of the EORTC change scores for “deterioration” at t2 (body image, speech, and physical functioning) and for seven EORTC change scores for “improvement” at t3 (social eating, swallowing, emotional functioning, social functioning, fatigue, appetite loss, and the summary score); it was below our threshold of 0.15 in 13 scales for deterioration and in nine scales for improvement, suggesting that overall quality of life may not be a useful anchor for many of the EORTC scales (see eTable 4 for details).

All MIC_{mean} and MIC_{predict} are displayed in Table 4. Averaging together results from the MIC_{mean} approach (evaluating average change scores among the subset of patients who reported “a little” change on the SSQ) and MIC_{predict} approach (which used predictive modelling to predict broader levels of change on the SSQ in the whole sample), the MIC for swallowing was 13 for deterioration and 8 for improvement; for speech it was 14 for deterioration (no average was computed for improvement because of poor correlation); for dry mouth it was 26 for deterioration and 6 for improvement; and for global quality of life it was 10 for deterioration and 5 for improvement. Generally, the MICs based on specific anchors differentiated well between the patient groups (Figure 2).

Table 2
Number and proportion of patients who reported their swallowing, speech, dry mouth and overall quality of life to be deteriorated between baseline (t1) and 3 months later (t2) and improved between 3 (t2) and 6 months (t3) after baseline.

	Between t1 and t2		Between t2 and t3	
	A little worse*	Deteriorated* *	A little better*	Improved* *
Swallowing	34 (16 %)	100 (47 %)	42 (21 %)	108 (50 %)
Speech	36 (18 %)	89 (42 %)	34 (17 %)	77 (36 %)
Dry Mouth	47 (22 %)	135 (64 %)	42 (22 %)	65 (31 %)
Overall Quality of Life	40 (19 %)	107 (50 %)	58 (29 %)	109 (51 %)

Note: Percentages were ascertained using the Subjective Significance Questionnaire (SSQ).

* The columns “A little worse” and “A little better” display the subgroups of patients who said their quality of life had changed a little.

* * The columns “Deteriorated” and “Improved” combine the different categories of worsening and improvement. That is, they include all patients who said their quality of life in the respective domain was very much changed, moderately changed, or a little changed.

Table 3
Minimal detectable difference (MDD) estimates.

EORTC Scale	t1		t2		t3		Av
<i>EORTC QLQ-HN43</i>	SEM	0.5 SD	SEM	0.5 SD	SEM	0.5 SD	
Anxiety	13	16	13	16	14	15	14
Body Image	11	11	12	14	11	14	12
Dry mouth and sticky saliva	14	12	15	16	16	15	15
Pain in the mouth	12	11	11	13	10	11	11
Sexuality	11	16	13	19	11	18	15
Senses	11	11	19	15	17	14	14
Shoulder problems	9	9	10	13	10	13	11
Skin problems	8	7	12	12	10	9	10
Social eating	9	12	12	15	11	14	12
Speech	9	13	10	15	10	13	12
Swallowing	9	11	11	14	10	13	11
Problems with teeth	13	13	15	13	14	13	14
Coughing	15	13	17	16	16	15	15
Swelling in the neck	17	16	18	16	17	16	17
Neurological problems	12	11	15	14	14	13	13
Opening mouth	14	13	17	16	16	14	15
Social contact	14	13	20	18	17	15	16
Weight loss	15	14	19	17	17	16	16
Wound healing	12	11	14	13	13	12	12
<i>EORTC QLQ-C30</i>							
Physical functioning	8	9	16	12	21	11	13
Role functioning	26	14	26	17	28	15	21
Emotional functioning	22	12	22	13	22	12	17
Cognitive functioning	15	11	15	11	18	11	14
Social functioning	26	14	26	16	27	15	20
Global health status/ QoL	19	11	19	11	20	11	15
Fatigue	23	12	23	14	24	13	18
Nausea and vomiting	8	6	8	11	15	8	9
Pain	25	14	25	14	24	13	19
Dyspnoea	22	13	22	14	21	13	17
Insomnia	26	16	26	17	25	15	21
Appetite loss	25	15	25	18	26	15	21
Constipation	22	13	22	15	25	15	19
Diarrhoea	14	8	14	12	17	10	13
Financial difficulties	24	14	24	16	25	15	20
Summary Score	15	8	15	9	17	9	12

Note: SEM = Standard error of the mean, SD = Standard deviation, Av = Average of all estimates

Secondary analyses used the *overall quality of life* SSQ item as a global anchor to explore the MIC for all other EORTC scales. Among the 35 EORTC scales, we were able to average together results from the MIC_{mean} and MIC_{predict} approaches for 22 scales for deterioration and for 27 scales for improvement (correlations between EORTC change scores and the global anchor were below our threshold for the remaining scales). In these ancillary analyses, the average MIC for deterioration ranged from 1, for *swelling in the neck*, to 30, for *dry mouth* (when using the global anchor). The average MIC for improvement ranged from 4, for *insomnia*, to 18, for *appetite loss*. Not surprisingly, use of the global anchor often resulted in higher MIC estimates than use of the specific anchors, though results were variable. As noted, many of the correlations with the global anchor were poor, and this approach was considered suboptimal.

3.6. Minimal detectable change (MDC)

The minimal detectable EORTC change scores, taken as the average of SEM and SD-based estimates, for deterioration ranged from 9, both for *physical functioning* and for the *summary score*, to 24, for *appetite loss*. For improvement, it ranged from 8, both for *physical functioning* and for the *summary score*, to 22, for *appetite loss* (Table 5).

4. Discussion

With this study, we aimed to establish interpretation thresholds for the EORTC QLQ-C30 and the EORTC QLQ-HN43 used for patients with head and neck cancer (Table 6). These questionnaires are among the most commonly employed measures of HRQoL among head and neck cancer patients around the world [39]. There is keen interest among practitioners and investigators in developing thresholds to help interpret scores, and in identifying the minimal differences or changes in scores that are clinically meaningful. To increase the rigor of our investigation, we drew on a number of different methodological approaches to explore the responses of a large cohort of head and neck cancer patients from 15 different countries, assessed over the course of six months.

We examined the MIC (i.e., minimal changes in EORTC scores within individuals over time) that were important to patients themselves, using an anchor (patient-rated changes) and employing a number of statistical methods (MIC_{mean} and MIC_{predict}). A main finding is that, as shown by others [21], MIC estimates for improvements over time were generally smaller than the respective MIC estimates for deteriorations. This implies that patients often can accept certain declines in their HRQoL and are sensitive to or appreciate even small improvements. In our primary analyses, which used specific, conceptually-relevant anchors to examine changes in four of the most salient EORTC domains - swallowing, speech, dry mouth, and global quality of life - the MIC values ranged from 5 to 8 for improvement over time, and from 10 to 26 for deterioration.

In ancillary analyses, we explored the value of using a global or less direct anchor, patient ratings of changes in overall QoL, to generate MIC estimates for each of the other EORTC scales (again examining MIC estimates for improvement over time separately from estimates for deterioration, and using two different statistical methods). The MIC values were highly variable (ranging from 1 to 30), and many of the correlations between the EORTC change scores and the anchor were small, suggesting that this approach is not optimal.

Another important aspect of our findings is that, for declines over time in global quality of life, the often used 10-point-rule of thumb [35] is backed by our data. However, one should keep in mind here again that for improvement, the estimate is smaller: an improvement of 5 points can already be considered to be clinically relevant (using change in a specific, conceptually relevant anchor as the criterion). For other EORTC scales where we have a specific anchor available (i.e., swallowing, speech, dry mouth), the commonly used 10-point estimate would be lower than our MIC values for deterioration and would exceed our MIC values for improvement. Similar results were recently found for another frequently used instrument, especially in human papillomavirus-associated oropharyngeal cancer trials, the MD Anderson Dysphagia Inventory (MDADI), where estimates smaller than the usual 10-point-rule of thumb turned out to be of clinical relevance [40]. If the MIC is used for sample size calculation for a clinical trial, the consequences are considerable.

Moreover, as noted by others [21,40,41], a single number is not likely to offer a universal threshold for identifying minimal important changes on diverse HRQoL scales. Another point to consider is that the MICs might differ between patient groups with different cancer locations in the head and neck area or receiving different treatments. This could be explored in future studies.

Musoro et al. [38] generally found different MIC values for selected EORTC QLQ-C30 scales among patients with head and neck cancer than we did. Their analyses were based on clinical trial data and using performance status or CTCAE criteria as an anchor, and using the MIC_{mean} as method. For example, for global health status/quality of life they reported 9 points as the MIC for improvement and 5 points for deterioration (thus here, the estimate for deterioration was lower than for improvement, which is in contrast to our results and to those of other studies [21]). Their MIC for improved dyspnea was 6, which is the same estimate as in our study. However, their MIC for improved social

Table 4
Minimal important change (MIC) estimates.

EORTC Scale		Deterioration				Improvement			
		MIC _{mean}	MIC _{predict}	Corr	Av	MIC _{mean}	MIC _{predict}	Corr	Av
Specific Anchor	Swallowing	11	15	0.35	13	14	3	0.36	8
	Speech	15	13	0.40	14	8	5 *	0.15	
	Dry Mouth	25	27	0.49	26	11	1 *	0.24	6
	Global QoL	10	10	0.30	10	2	8	0.42	5
Global Anchor	Anxiety	8	7	0.13		14	3	0.15	8
	Body Image	16	13	0.39	15	12	2	0.23	7
	Dry mouth	30	30	0.28	30	5	4	0.22	5
	Pain in the mouth	7	7	0.21	7	15	5	0.15	10
	Sexuality	11	7	0.19	9	8	11	0.21	9
	Senses	22	25	0.14		16	6	0.27	11
	Shoulder problems	12	16	0.02		3	131	0.01	
	Skin problems	13	15	0.20	14	13	3	0.13	
	Social eating	13	22	0.23	18	15	3	0.33	9
	Speech	12	12	0.35	12	8	5	0.18	7
	Swallowing	8	12	0.20	10	15	3	0.33	9
	Problems with teeth	4	144	0.00		2	9	0.12	
	Coughing	7	2	0.18	4	13	4	0.29	8
	Swelling in the neck	2	0	0.19	1	11	11	0.19	11
	Neurological probl.	2	28	0.02		2	19	0.09	
	Opening mouth	15	12	0.20	13	7	26	0.03	
	Social contact	11	12	0.29	11	18	7	0.16	12
	Weight loss	14	12	0.15		10	9	0.10	
	Wound healing	1	5	0.22	3	6	9	0.18	8
	Physical functioning	17	14	0.34	16	10	5	0.22	7
	Role functioning	29	23	0.23	26	18	11	0.26	15
	Emotional function.	1	1	0.13		17	5	0.33	11
	Cognitive functioning	8	5	0.15	6	9	1	0.26	5
	Social functioning	17	15	0.27	16	20	8	0.34	14
	Global QoL	10	10	0.30	10	15	8	0.42	12
	Fatigue	24	17	0.28	21	15	8	0.32	11
	Nausea and vomiting	9	10	0.16	9	22	7	0.28	14
	Pain	6	5	0.13		17	7	0.18	12
	Dyspnoea	7	2	0.07		9	3	0.16	6
	Insomnia	3	5	0.10		8	1	0.17	4
	Appetite loss	13	20	0.13		25	12	0.30	18
	Constipation	2	8	0.08		2	18	0.10	
	Diarrhoea	10	9	0.14		14	2	0.20	8
	Financial difficulties	11	7	0.16	9	5	2	0.10	
	Summary Score	11	10	0.28	10	14	7	0.36	11

Note: Displayed are the absolute values of the minimal important change scores based on the mean approach (MIC_{mean}) and predictive modelling (MIC_{predict}), their average (Av), and the absolute values of the point biserial correlation (Corr) between the EORTC change score and the anchor from the Subjective Significance Questionnaire.

The primary analyses involved the specific anchors, whereas exploratory analyses used the global anchor.

Gray shading indicates insufficiently strong association (threshold $r \leq 0.15$). In these cases, no average is presented.

* marks the MIC_{predict} values that were adjusted because of prevalence of improvement < 0.40 .

functioning was 7, which is lower than our MIC of 14. It is possible that their anchors (judgments by clinicians) are less suited than ours (the patients themselves judged their change) for assessing MIC in quality of life domains. Notably, many of the EORTC QLQ-C30 scales were excluded from the MIC analyses in that study due to lack of suitable anchors in the database, and some of the anchors that were retained correlated poorly with EORTC change scores. We encountered a similar problem in our secondary analyses, where the overall quality of life SSQ anchor often had only small correlations with the EORTC change scores.

Our study is the first to establish MIC estimates for the EORTC QLQ-HN43. It is therefore not possible to compare MIC for these scales with other studies. It can be noted though that there were substantial differences in MICs for the various scales. Hence, researchers need to decide carefully which domains they are most interested in and use the relevant MIC for that endpoint. This should also be considered when comparing HRQoL across trials.

From a methodological point of view, it is noteworthy that the MIC_{predict} was generally smaller than the MIC_{mean}. We decided to calculate an average of both, which is of course debatable. One could also decide to base their decision solely on the MIC_{predict} which has the

advantage of better power and perhaps less bias [5].

It should also be emphasised that specific anchors were indeed more suitable for calculating the MIC than the global or indirect anchor “overall quality of life”, as evidenced by the poor correlation of this anchor with many of the change scores. We had already suspected this before the analysis, but had to balance the burden on patients with the value of obtaining better anchors. For future studies, however, the importance of using specific, conceptually-relevant anchors that are well correlated with the HRQoL indices should be emphasized. This is a clear limitation of our study, and a salient challenge for further investigations.

Aside from the MIC, this study evaluated a number of other important thresholds for interpreting EORTC scores (MID, MDD MDC). Our efforts to use performance status as an anchor to examine the MID (i.e., the minimal important difference between groups) were not successful. Consistent with some prior investigations [16,28], clinician-rated performance status was not strongly correlated with patient-reported HRQoL scores. This is not unanticipated, since HRQoL instruments were developed as an urgently needed method of moving away from reliance on other people’s evaluations and emphasizing instead patients’ appraisals of their own experience. Clinician-rated performance status

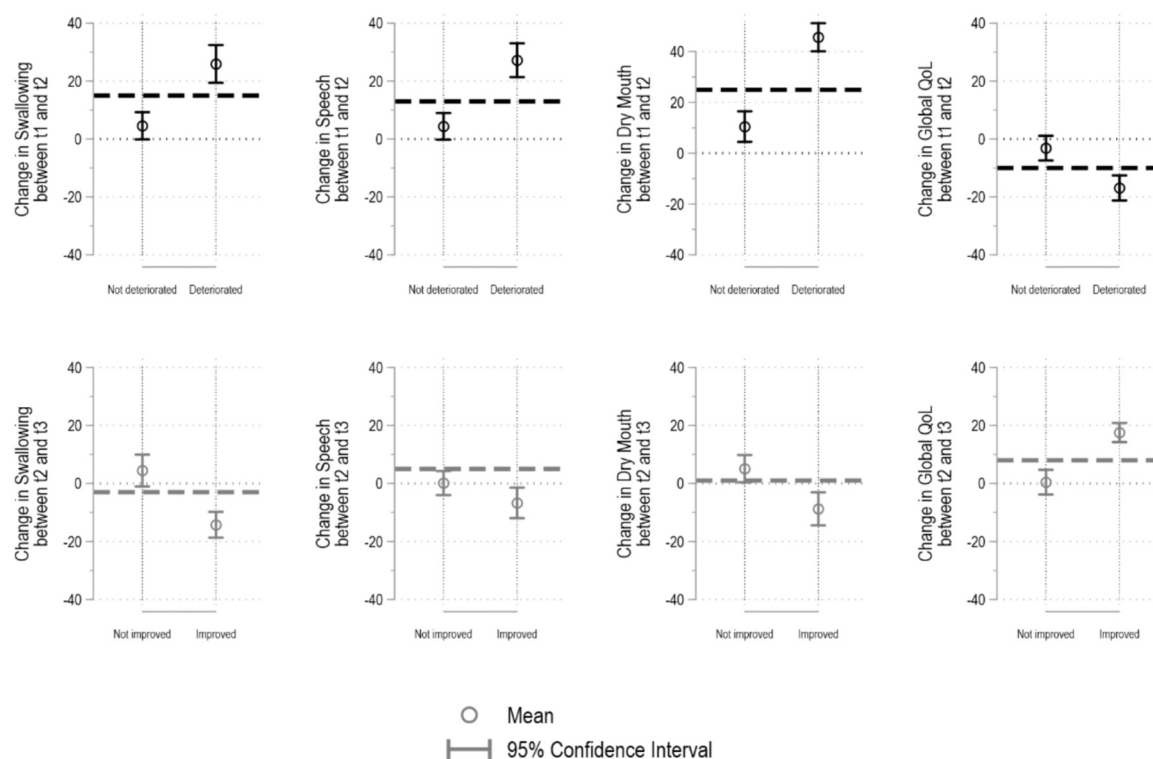


Fig. 2. Distribution of EORTC change scores per Subjective Significance Questionnaire group (deteriorated vs. no change and improved vs. no change). *Notes:* The EORTC change scores represent the difference in the score for the particular quality of life scale from the current time point compared to the previous one. Change scores > 0 in the swallowing, speech, and dry mouth scales and < 0 in global quality of life represent a deterioration of quality of life. Change scores < 0 in the swallowing, speech, and dry mouth scales and > 0 in global quality of life represent an improvement in quality of life. The dashed lines represent the minimal important change score based on predictive modelling (MIC_{predict}).

may be a convenient metric, but not always a very useful one for generating criteria with which to evaluate differences in HRQoL scores.

Minimal detectable differences and change scores, which use distribution-based methods to calculate benchmarks based on the level of statistical certainty, sometimes have been considered to be less optimal than anchor-based MIDs and MICs by some investigators, but they still provide useful information [2]. We evaluated the MDD using distribution-based approaches. Drawing on results from two commonly used statistical methods (SEM and 0.5 SD) at three points in time, we found that the MDD values for the EORTC QLQ-HN43 scales ranged from 10 to 17 (average = 14). The MDDs for the EORTC QLQ-C30 core module scales were generally somewhat higher, ranging from 9 to 21 (average = 17). Next to the 0.5 SD, it is also common to use 0.2 or 0.3 SD as a distribution-based method. We had decided to only use the 0.5 SD because it had worked well in our pilot study [28] and because it is more conservative.

Using a similar approach to evaluate the MDC (i.e., the minimal detectable change over time), we found MDC estimates that ranged from 11 to 19 for deterioration and from 11 to 17 for improvement on the EORTC QLQ-HN43 scales, and from 9 to 24 for deterioration and from 8 to 22 for improvement on the EORTC QLQ-C30 scales. For our primary outcomes (swallowing, speech, dry mouth, and global quality of life), these estimates generally fell between the MIC values for improvement and deterioration.

A previous study [38] also evaluated the MDD for several EORTC QLQ-C30 scales, though no studies have examined these metrics for the EORTC QLQ-HN43 scales. At the start of treatment, Musoro et al. [41] found, for example, for physical functioning the SEM-based MDD to be 5 (this was 8 in our study before treatment) and SD-based MDD to be 9 (also 9 in our study). There are, of course, differences in estimates generated from different scales and at different time points. In our study, the MDD and MDC were generally higher than the MIC (Table 6).

4.1. Strengths and limitations

This investigation addresses an important gap in the literature by examining for the first time important changes and differences for interpreting the revised EORTC QLQ-HN43 module. The study evaluated MIC thresholds as well as MDD and MDC estimates using multiple approaches. Other strengths include the large, multinational sample and the longitudinal research design, encompassing three assessment points. The assessment periods appear to have been well chosen because the proportion of patients classified as improved and deteriorated was well balanced, which contributes to valid results [37].

Important limitations include lack of a suitable anchor to examine MIDs; the fact that clinician-rated performance status was not a very useful criterion contributes to the literature regarding the utility of different commonly available anchors. However, defining more suitable anchors is not easy. One solution could be to collect more information on supportive care and treatment changes (because this is how the MID is defined). This could work well for *Pain*, for example, but likely not for domains where no treatment is possible, or at least no “simple” treatment (e.g., *Fatigue*, *Body Image*). An additional limitation is that specific anchors were available to examine the MIC for four EORTC scales, whereas only an indirect global anchor could be used to evaluate the remaining scales. The more limited performance of the global anchor is a notable finding that may contribute to future research in this area, by encouraging use of specific, conceptually related anchors—ideally those evaluated by patients. Further, the sample sizes per SSQ category are small which is particularly relevant for the MIC_{mean} as it limits the statistical certainty for these estimates.

4.2. Application in research and clinic

Both practitioners and researchers can make use of the values that we

Table 5
Minimal detectable change (MDC) estimates.

EORTC Scale	Deterioration				Improvement			
	ICC	SEM	0.5 SD	Av	SEM	0.5 SD	Av	
<i>EORTC QLQ-HN43</i>								
Anxiety	0.78	17	18	18	13	18	16	
Body Image	0.88	9	14	12	8	14	11	
Dry mouth and sticky saliva	0.87	12	17	15	11	17	14	
Pain in the mouth	0.88	13	18	15	9	18	14	
Sexuality	0.91	12	20	16	11	20	16	
Senses	0.87	10	14	12	10	14	12	
Shoulder problems	0.95	8	17	12	7	17	12	
Skin problems	0.73	15	14	15	12	14	13	
Social eating	0.93	10	19	14	8	19	13	
Speech	0.89	12	17	14	11	17	14	
Swallowing	0.86	12	17	15	10	17	13	
Problems with teeth	0.89	9	14	11	7	14	10	
Coughing	0.75	13	13	13	11	13	12	
Swelling in the neck	0.52	23	16	19	18	16	17	
Neurological problems	0.88	10	15	13	8	15	12	
Opening mouth	0.87	11	16	13	9	16	12	
Social contact	0.62	17	14	15	13	14	13	
Weight loss	0.87	12	17	15	10	17	14	
Wound healing	0.73	16	15	16	14	15	15	
<i>EORTC QLQ-C30</i>								
Physical functioning	0.87	7	10	9	7	10	8	
Role functioning	0.83	15	18	17	13	18	16	
Emotional functioning	0.81	11	13	12	10	13	12	
Cognitive functioning	0.86	8	11	10	7	11	9	
Social functioning	0.81	15	17	16	13	17	15	
Global health status/ QoL	0.81	11	12	11	9	12	11	
Fatigue	0.81	12	14	13	11	14	13	
Nausea and vomiting	0.45	17	12	14	18	12	15	
Pain	0.84	13	16	14	11	16	13	
Dyspnoea	0.83	12	15	13	10	15	12	
Insomnia	0.77	18	19	19	17	19	18	
Appetite loss	0.65	26	22	24	22	22	22	
Constipation	0.86	13	18	15	13	18	15	
Diarrhoea	0.68	16	14	15	15	14	14	
Financial difficulties	0.84	11	14	12	10	14	12	
Summary Score	0.78	9	9	9	8	9	8	

Note: SEM = Standard error of the mean, SD = Standard deviation. Av = Average of SEM- and SD-based MDC estimates

derived. They should keep in mind that the estimates serve different purposes: MIC and MDC are benchmarks for the interpretation of change scores (changes within an individual); MID and MDD are benchmarks for the interpretation of differences between groups.

The MIC can be used for different purposes: a) as a threshold to determine the number of responders in clinical trials. This may add the patients' perspective to other outcomes. b) as a threshold to determine the number of responders in groups of patients receiving a particular treatment in order to inform patients about the expected effect of the treatment, which can facilitate shared decision-making. The MIC can be used not only as a (deterministic) threshold for response, but also as a probabilistic value to interpret the change scores in terms of the likelihood that an individual patient has experienced a significant change. The MDC is also a relevant benchmark for change but it lacks the patient perspective. MID and MDD can be used to interpret score differences between patient groups in clinical studies.

It should be borne in mind though that the estimates are based on an average of patients. While the percentage of responders at the group level is likely to be correct, each patient may have their own MIC, and therefore some individuals may be misclassified. In addition, measurement errors contribute to the misclassification of individual patients.

5. Conclusions

Our results underline that no single MIC or MDC can be applied to all EORTC QLQ-HN43 scales, and that the MIC for deterioration seems larger than those for improvement.

Table 6
Overview of all average values of MIC, MDC, and MDD.

EORTC Scale		MIC	MIC	MDC	MDC	MDD
		Deter.	Improv.	Deter.	Improv.	
Specific Anchor	Swallowing	13	8			
	Speech	14				
	Dry Mouth	26	6			
	Global Quality of Life	10	5			
Global Anchor	<i>EORTC QLQ-HN43</i>					
	Anxiety		8	18	16	14
	Body Image	15	7	12	11	12
	Dry mouth and sticky saliva	30	5	15	14	15
	Pain in the mouth	7	10	15	14	11
	Sexuality	9	9	16	16	15
	Senses		11	12	12	14
	Shoulder problems			12	12	11
	Skin problems	14		15	13	10
	Social eating	18	9	14	13	12
	Speech	12	7	14	14	12
	Swallowing	10	9	15	13	11
	Problems with teeth			11	10	14
	Coughing	4	8	13	12	15
	Swelling in the neck	1	11	19	17	17
	Neurological problems			13	12	13
	Opening mouth	13		13	12	15
	Social contact	11	12	15	13	16
	Weight loss			15	14	16
	Wound healing	3	8	16	15	12
	<i>EORTC QLQ-C30</i>					
	Physical functioning	16	7	9	8	13
	Role functioning	26	15	17	16	21
	Emotional functioning		11	12	12	17
	Cognitive functioning	6	5	10	9	14
	Social functioning	16	14	16	15	20
	Global health status/QoL	10	12	11	11	15
	Fatigue	21	11	13	13	18
	Nausea and vomiting	9	14	14	15	9
	Pain		12	14	13	19
	Dyspnoea		6	13	12	17
	Insomnia		4	19	18	21
	Appetite loss		18	24	22	21
	Constipation			15	15	19
	Diarrhoea		8	15	14	13
	Financial difficulties	9		12	12	20
	Summary Score	10	11	9	8	12

Deter = Deterioration, Improv = Improvement; MIC = Minimal important change, MDC = Minimal detectable change, MDD = Minimal detectable difference.

This table combines the average scores from Tables 3, 4, and 5.

The findings can help researchers and clinicians to interpret the scores of the EORTC QLQ-C30 and the EORTC QLQ-HN43 in patients with head and neck cancer and to plan clinical trials. We were able to use specific anchors for four domains, and we saw that specific anchors were indeed superior to a global anchor in calculating MICs. Therefore, future research should use specific anchors whenever possible.

Ethics approval

All sites obtained ethical approval in accordance with regional and national requirements. Approval number from the principal investigator's institution: 837.281.14–9520.

Consent to participate

Patients were given time to consider the study and ask any questions before consenting and participating. All participants provided written informed consent.

Consent for publication

Not applicable.

Funding

This work was supported by the EORTC Quality of Life Group [grant number 001/2014] and partially supported by the Japan Agency for Medical Research and Development (AMED) [grant number JP19ck0106498].

CRedit authorship contribution statement

Marcos Santos: Writing – review & editing, Investigation. **Joaquim Castro Silva:** Writing – review & editing, Investigation. **Hisham Mehanna:** Writing – review & editing, Investigation, Conceptualization. **Monica Pinto:** Writing – review & editing, Investigation, Conceptualization. **Juan Ignacio Arraras:** Writing – review & editing, Investigation. **Noam Yarom:** Writing – review & editing, Investigation, Conceptualization. **Pierluigi Bonomo:** Writing – review & editing, Investigation, Conceptualization. **Amy Fullerton:** Writing – review & editing, Investigation, Conceptualization. **Loreto Fernandez Gonzalez:** Writing – review & editing, Investigation, Conceptualization. **Johanna Inhestern:** Writing – review & editing, Investigation, Conceptualization. **Allen C. Sherman:** Writing – review & editing, Investigation, Conceptualization. **Ingo Baumann:** Writing – review & editing, Investigation. **Lisa Licitra:** Writing – review & editing, Conceptualization. **Jean-Pascal Machiels:** Writing – review & editing, Conceptualization. **Kristin Bjordal:** Writing – review & editing, Supervision, Conceptualization. **Razvan Galalae:** Writing – review & editing, Investigation. **Ouranica Nicolatou-Galitis:** Writing – review & editing, Investigation, Conceptualization. **Andreas Boehm:** Writing – review & editing, Investigation, Conceptualization. **Eva Hammerlid:** Writing – review & editing, Investigation, Conceptualization. **Sanja Krejovic-Trivic:** Writing – review & editing, Investigation. **Iwona M. Tomaszewska:** Writing – review & editing, Investigation. **Wei-Chu Chie:** Writing – review & editing, Investigation. **Cecilie Delphin Amdal:** Writing – review & editing, Investigation, Conceptualization. **Katherine J. Taylor:** Writing – review & editing, Investigation, Conceptualization. **Bente Brokstad Herlofson:** Writing – review & editing, Investigation, Conceptualization. **Naomi Kiyota:** Writing – review & editing, Investigation, Conceptualization. **Judith Raber-Durlacher:** Writing – review & editing, Investigation, Conceptualization. **Dina Salem:** Writing – review & editing, Investigation. **Alexander Fabian:** Writing – review & editing, Investigation, Conceptualization. **Susanne Singer:** Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

Authors declare they have no conflicts of interest.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.ejca.2024.115062](https://doi.org/10.1016/j.ejca.2024.115062).

Data Availability

The data of this study are stored in the EORTC data repository and can be accessed by other researchers.

References

- [1] King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharm Outcomes Res* 2011;11(2):171–84.
- [2] Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003;56(5):395–407.
- [3] Jaeschke R, Singer J, Guyatt GH. Measurement of health-status - ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10(4):407–15.
- [4] de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes* 2006;4:5.
- [5] Terwee CB, Peipert JD, Chapman R, Lai JS, Terluin B, Cella D, et al. Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. *Qual Life Res* 2021;30(10):2729–54.
- [6] Terluin B, Eekhout I, Terwee C, De Vet HCW. Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. *J Clin Epidemiol* 2015;68:1388–96.
- [7] Norman GR, Sloan JA, Wywich KW. Interpretation of changes in health-related quality of life - the remarkable universality of half a standard deviation. *Med Care* 2003;41(5):582–92.
- [8] Aaronson N, Ahmedzai S, Bergmann B, Bullinger M, Cull A, Duez NJ, et al. The European organization for research and treatment of cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Nat Cancer Inst* 1993;85(5):365–76.
- [9] Bjordal K, de GA, Fayers PM, Hammerlid E, van PC, Curran D, et al. A 12 country field study of the EORTC QLQ-C30 (version 3.0) and the head and neck cancer specific module (EORTC QLQ-H&N35) in head and neck patients. *EORTC Quality of Life Group. Eur J Cancer* 2000;36(14):1796–807.
- [10] Singer S, Amdal CD, Hammerlid E, Tomaszewska IM, Silva JC, Mehanna H, et al. International validation of the revised European Organisation for Research and Treatment of Cancer Head and Neck Cancer Module, the EORTC QLQ-HN43: phase IV. *Head Neck* 2019;41(6):1725–37.
- [11] Weis J, Tomaszewski K, Hammerlid E, Arraras J, Conroy T, Lancelley A, et al. International psychometric validation of an EORTC quality of life module measuring cancer related fatigue (EORTC QLQ-FA12). *J Natl Cancer Inst* 2017;109(5). Article number djw273.
- [12] Musoro J, Coens C, Fiteni F, Katarzyna P, Cardoso F, Russell NS, et al. Minimally important differences for interpreting EORTC QLQ-C30 scores in patients with advanced breast cancer. *JNCI Cancer Spectr* 2019;3(3):pkz037.
- [13] Raman S, Ding KY, Chow E, Meyer RM, van der Linden YM, Roos D, et al. Minimal clinically important differences in the EORTC QLQ-C30 and brief pain inventory in patients undergoing re-irradiation for painful bone metastases. *Qual Life Res* 2018;27(4):1089–98.
- [14] Koller M, Musoro JZ, Tomaszewski K, Coens C, King MT, Sprangers MAG, 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;167:65–72.
- [15] Dirven L, Musoro JZ, Coens C, Reijneveld JC, Taphoorn MJB, Boele FW, et al. Establishing anchor-based minimally important differences for the EORTC QLQ-C30 in glioma patients. *Neuro-Oncol* 2021;23(8):1327–36.
- [16] Kawahara T, Taira N, Shiroiwa T, Hagiwara Y, Fukuda T, Uemura Y, 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.
- [17] Maringwa JT, Quinten C, King M, Ringash J, Osoba D, Coens C, et al. 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.
- [18] Quinten C, Kenis C, Decoster L, Debruyne PR, De Groof I, Focan C, et al. Determining clinically important differences in health-related quality of life in older patients with cancer undergoing chemotherapy or surgery. *Qual Life Res* 2019;28(3):663–76.
- [19] Musoro J, Sodergren SC, Coens C, Pochesci A, Terada M, King MT, 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.
- [20] Musoro J, Coens C, Sprangers MAG, Brandberg Y, Groenvold M, Flechtner H, et al. Minimally important differences for interpreting EORTC QLQ-C30 change scores over time: a synthesis across 21 clinical trials involving nine different cancer types. *Eur J Cancer* 2023;188:171–82.
- [21] Cocks K, King MT, Velikova G, de Castro G, St James MM, Fayers PM, et al. Evidence-based guidelines for interpreting change scores for the European

- organisation for the research and treatment of cancer quality of life questionnaire core 30. *Eur J Cancer* 2012;48(11):1713–21.
- [22] Clarke NA, Braverman J, Worthy G, Shaw JW, Bennett B, Dhanda D, et al. A review of meaningful change thresholds for EORTC QLQ-C30 and FACT-G within oncology. *Value Health* 2024;27(4):458–68.
- [23] Sully K, Trigg A, Bonner N, Moreno-Koehler A, Trenner C, Shah N, et al. Estimation of minimally important differences and responder definitions for EORTC QLQ-MY20 scores in multiple myeloma patients. *Eur J Haematol* 2019;103(5):500–9.
- [24] Reni M, Braverman J, Hendifar A, Li CP, Macarulla T, Oh DY, et al. Evaluation of minimal important difference and responder definition in the EORTC QLQ-PAN26 module for assessing health-related quality of life in patients with surgically resected pancreatic adenocarcinoma. *Ann Surg Oncol* 2021;28(12):7545–54.
- [25] Coomans MB, Taphoorn MJB, Aaronson NK, Baumert BG, van den Bent M, Bottomley A, et al. Measuring change in health-related quality of life: the impact of different analytical methods on the interpretation of treatment effects in glioma patients. *Neuro-Oncol Pr* 2020;7(6):668–75.
- [26] Singer S, Araújo C, Arraras J, Baumann I, Boehm A, Herlofson BB, et al. Measuring quality of life in head and neck cancer patients - update of the EORTC QLQ-H&N module, phase III. *Head Neck* 2015;37(9):1358–67.
- [27] Singer S, Arraras J, Baumann I, Boehm A, Chie WC, Galalae R, et al. Quality of life in head and neck cancer patients receiving multimodal or targeted therapy - update of the EORTC QLQ-H&N35, Phase I. *Head Neck* 2013;35(9):1331–8.
- [28] Singer S, Hammerlid E, Tomaszewska IM, Amdal CD, Bjordal K, Herlofson BB, et al. Methodological approach for determining the Minimal Important Difference and Minimal Important Change scores for the European Organisation for Research and Treatment of Cancer Head and Neck Cancer Module (EORTC QLQ-HN43) exemplified by the Swallowing scale. *Qual Life Res* 2022;31(3):841–53.
- [29] Kvam AK, Fayers P, Wisloff F. What changes in health-related quality of life matter to multiple myeloma patients? A prospective study. *Eur J Haematol* 2010;84(4):345–53.
- [30] Terluin B, Brouwers EPM, Marchand MAG, de Vet HCW. Assessing the equivalence of Web-based and paper-and-pencil questionnaires using differential item and test functioning (DIF and DTF) analysis: a case of the Four-Dimensional Symptom Questionnaire (4DSQ). *Qual Life Res* 2018;27(5):1191–200.
- [31] Vach W, Saxer F. Anchor-based minimal important difference values are often sensitive to the distribution of the change score. *Qual Life Res* 2024;33(5):1223–32.
- [32] Singer S, Langendijk JA, Yarom N. Assessing and improving quality of life in patients with head and neck cancer. *ASCO Educ Book* 2013:230–5.
- [33] Husson O, de Rooij BH, Kieffer J, Oerlemans S, Mols F, Aaronson NK, et al. The EORTC QLQ-C30 summary score as prognostic factor for survival of patients with cancer in the "real-world": results from the population-based PROFILES registry. *Oncologist* 2020;25(4):E722–32.
- [34] Giesinger JM, Kieffer JM, Fayers PM, Groenvold M, Petersen MA, Scott NW, et al. Replication and validation of higher order models demonstrated that a summary score for the EORTC QLQ-C30 is robust. *J Clin Epidemiol* 2016;69:79–88.
- [35] 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;16:139–44.
- [36] Devji T, Carrasco-Labra A, Qasim A, Phillips M, Johnston BC, Devasenapathy N, et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. *Br Med J* 2020;369. Article Number: m1714.
- [37] Terluin B, Eekhout I, Terwee CB. The anchor-based minimal important change, based on receiver operating characteristic analysis or predictive modeling, may need to be adjusted for the proportion of improved patients. *J Clin Epidemiol* 2017;83:90–100.
- [38] Griffiths P, Sims J, Williams A, Williamson N, Cella D, Brohan E, et al. How strong should my anchor be for estimating group and individual level meaningful change? A simulation study assessing anchor correlation strength and the impact of sample size, distribution of change scores and methodology on establishing a true meaningful change threshold. *Qual Life Res* 2023;32(5):1255–64.
- [39] Mehanna H, Carter B, Hartley A, Abou-Foul AK, Brooks J, Jones J, et al. Patient preference for commonly-used, head and neck cancer-specific quality of life questionnaires in the follow-up setting (Determin): a multi-centre randomised controlled trial and mixed methods study. *Clin Otolaryngol* 2023;48(4):613–22.
- [40] McDowell L, King MT, Hutcheson KA, Ringash J, Yom SS, Corry J, et al. A hard truth to swallow: critically evaluating the MD anderson dysphagia inventory (MDADI) as an endpoint in human papillomavirus-associated oropharyngeal cancer trials. *Int J Radiat Oncol Biol Phys* 2024. <https://doi.org/10.1016/j.ijrobp.2024.05.005>.
- [41] Musoro J, Coens C, Singer S, Tribius S, Oosting S, Russell N, et al. Minimally important differences for interpreting European Organisation for Research and Treatment for Cancer Quality of life Core Questionnaire scores in patients with head and neck cancer. *Head Neck* 2020;42(11):3141–52.