Original Study

Health-related Quality of Life in the Phase III LUME-Colon 1 Study: Comparison and Interpretation of Results From EORTC QLQ-C30 Analyses

Heinz-Josef Lenz,[1] Guillem Argiles,[2] Takayuki Yoshino,[3] Sara Lonardi,[4] Alfredo Falcone,[5] María Luisa Limón,[6] Alberto Sobrero,[7] Claudia Hastedt,[8] Barbara Peil,[9] Florian Voss,[9] Ingolf Griebsch,[10] Eric Van Cutsem[11]

Abstract

Based on European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30) outputs from the LUME-Colon 1 study, we compared and discussed different statistical methods for evaluating health-related quality of life data in oncology clinical trials. The different analyses consistently showed that patients' overall global health status/quality of life status was not impaired by active treatment with nintedanib versus placebo, and that patients perceived some benefits with nintedanib compared with placebo.

Introduction: We used European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30) data from the LUME-Colon 1 study to illustrate different methods of statistical analysis for health-related quality of life (HRQoL), and compared the results. **Patients and Methods:** Patients were randomized 1:1 to receive nintedanib 200 mg twice daily plus best supportive care (n = 386) or matched placebo plus best supportive care (n = 382). Five methods (mean treatment difference averaged over time, using a mixed-effects growth curve model; mixed-effects models for repeated measurements (MMRM); time-to-deterioration (TTD); status change; and responder analysis) were used to analyze EORTC QLQ-C30 global health status (GHS)/QoL and scores from functional scales. **Results:** Overall, GHS/QoL and physical functioning deteriorated over time. Mean treatment difference slightly favored nintedanib over placebo for physical functioning (adjusted mean, 2.66; 95% confidence interval [CI], 0.97-4.34) and social functioning (adjusted mean, 2.62; 95% CI, 0.66-4.47). GHS/QoL was numerically better with nintedanib versus placebo (adjusted mean, 1.61; 95% CI, −0.004 to 3.27). MMRM analysis had similar results, with better physical functioning in the nintedanib group at all timepoints. There was no significant delay in GHS/QoL deterioration (10%) and physical functioning (16%) with nintedanib versus placebo (TTD analysis). Status change analysis showed a higher proportion of patients with markedly improved GHS/QoL and physical functioning in the nintedanib versus placebo groups. Responder analysis showed a similar, less pronounced pattern. **Conclusion:** Analyses of EORTC QLQ-C30 data showed that HRQoL was not impaired by treatment with nintedanib versus placebo. Analysis and interpretation of HRQoL endpoints should consider symptom type and severity and course of disease.

Clinical Colorectal Cancer, Vol. 18, No. 4, 269-79

Keywords: HRQoL, MMRM, Nintedanib, QoL, Time to deterioration

[1]Division of Medical Oncology, USC Norris Comprehensive Cancer Center, Los Angeles, CA
[2]Medical Oncology Department, Vall d'Hebron University Hospital and Institute of Oncology (VHIO), Universitat Autònoma de Barcelona, Barcelona, Spain
[3]Department of Gastroenterology and Gastrointestinal Oncology, National Cancer Center Hospital East, Chiba, Japan
[4]Phase 1 Trial Unit and Medical Oncology Unit 1, Istituto Oncologico Veneto IRCCS, Padova, Italy
[5]Department of Oncology, University of Pisa, Pisa, Italy
[6]Medical Oncology Department, Hospital Universitario Virgen del Rocio, Sevilla, Spain
[7]Medical Oncology 1, Ospedale Policlinico San Martino IRCCS, Genova, Italy
[8]TA CNS Retinopathies Emerging Areas, Boehringer Ingelheim International GmbH, Ingelheim, Germany
[9]Biostatistics and Data Sciences, Boehringer Ingelheim Pharma GmbH & Co. KG, Ingelheim, Germany
[10]TA Oncology, Boehringer Ingelheim International GmbH, Ingelheim, Germany
[11]Digestive Oncology, University Hospitals Leuven and KULeuven, Leuven, Belgium

Submitted: Nov 26, 2018; Revised: Aug 1, 2019; Accepted: Aug 27, 2019; Epub: Sep 4, 2019

Address for correspondence: Heinz-Josef Lenz, MD, FACP, USC Norris Comprehensive Cancer Center, 1441 Eastlake Ave, NOR 3456, Los Angeles, CA 90089-9173
E-mail contact: lenz@usc.edu

https://doi.org/10.1016/j.clcc.2019.08.005

Introduction

Patient-reported outcomes (PROs) are an important consideration in the delivery of clinical care, clinical trial design, and comparative effectiveness research in oncology. Regulatory bodies such as the United States Food and Drug Administration and the European Medicines Agency encourage the inclusion of patient-focused clinical outcome assessments in clinical trials to support drug approvals and labeling claims.[1] PROs are also integrated into health technology assessments to reflect clinical benefit from the patient's perspective.[2]

In trials that include patients with advanced cancer, including those with metastatic colorectal cancer (CRC), the assessment and analysis of PROs present several challenges. Generally, there is an overall deterioration in patient health and quality of life (QoL) rather than an improvement, owing to the clinical course of the disease.[3] In addition, because of disease progression, there is a high dropout rate, which results in incomplete data sets.[3] The dropout may often differ between treatment arms, which complicates the issue of missing data.[4] Finally, health-related quality of life (HRQoL) scales are often multidimensional and include several subscales, which can lead to complex result patterns.[5]

The European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30) is often used to collect data on HRQoL and symptom burden in patients with cancer.[6] However, these data can be analyzed using a variety of methods, as exemplified in different clinical studies of CRC (see Supplemental Table 1 in the online version).[7-18] These statistical methods can be broadly allocated to 2 types: those that assess group means and those that assess clinically relevant changes at the level of the patient. The first type of method – group mean methods – treat the data as continuous endpoints, and include longitudinal models such as mixed-effects growth curve models and mixed-effects models for repeated measurements (MMRM). The second type of method requires definition of a minimal clinically important difference (MCID) at the level of the individual patient, including time-to-deterioration (TTD), distribution of patients who improved, patients who were stable or worsened from baseline, and responder analysis.[19]

Nintedanib is an oral, twice-daily (bid), triple angiokinase inhibitor that targets vascular endothelial growth factor receptors 1 to 3, platelet-derived growth factor receptors α/β, and fibroblast growth factor receptors 1 to 3, as well as RET and Flt3 receptors.[20,21]

The randomized, double-blind, placebo-controlled, phase III LUME-Colon 1 study (ClinicalTrials.gov identifier NCT02149108) investigated the efficacy and safety of nintedanib plus best supportive care (BSC) versus placebo plus BSC for the treatment of patients with metastatic CRC refractory to standard therapies.[22] Co-primary endpoints were progression-free survival (PFS) and overall survival (OS).[22] PFS by independent central review was significantly longer with nintedanib than with placebo (hazard ratio [HR], 0.58; 95% confidence interval [CI], 0.49-0.69; $P < .0001$),[22] whereas OS was not (HR, 1.01; 95% CI, 0.86-1.19; $P = .8659$).[22] The safety profile of nintedanib was consistent with safety data previously reported in early-phase CRC studies.[23,24] In LUME-Colon 1, the most frequent grade ≥ 3 adverse events (AEs) in the nintedanib group were liver-related AEs, mainly increased alanine aminotransferase (8%) and aspartate aminotransferase (8%) levels and fatigue. The most common AEs leading to treatment discontinuation in the nintedanib group were fatigue (n = 8), followed by asthenia, decreased appetite, and malignant neoplasm progression (all n = 6).[22] The LUME-Colon 1 study incorporated assessments of PROs, including HRQoL, as a further endpoint.

The objectives of this evaluation were: (1) to perform different exploratory analyses of EORTC QLQ-C30 data from the LUME-Colon 1 study and (2) to discuss possible interpretations of these different evaluations and compare and contrast the various methods used.

Patients and Methods

Study Design

The study design of LUME-Colon 1 has been reported previously.[22] LUME-Colon 1 enrolled adults with histologically or cytologically confirmed metastatic or locally advanced CRC that was not amenable to curative surgery and/or radiotherapy. Patients were required to have progressed on approved standard therapies or to have experienced unacceptable toxicity.[22] Study participants were randomized 1:1 to receive treatment with either nintedanib 200 mg bid plus BSC (n = 386) or placebo bid plus BSC (n = 382) in 21-day cycles.[22] In this study, BSC was defined as the best palliative care, excluding specific anticancer treatments, according to the investigator's decision.

PRO Measures

PROs, including HRQoL, were evaluated using EORTC QLQ-C30 and the EuroQoL 5-dimensions questionnaire (EQ-5D). The EQ-5D was included for health economic analyses, and so the analyses in this report will only focus on the EORTC QLQ-C30 component. The EORTC QLQ-C30 consists of multi-item scales and single-item measures as follows: 5 functional scales (physical, role, emotional, cognitive, and social functioning)[25]; 3 symptom scales (fatigue, pain, and nausea/vomiting)[25]; 6 single questions (assessing dyspnea, appetite loss, sleep disturbance, constipation, diarrhea, and the perceived financial impact of disease and treatment)[25]; and global health status (GHS)/quality of life (QoL) scale.[25]

We report the GHS/QoL scale (consisting of 2 items) and the 5 functional scales (see Supplemental Table 2 in the online version).[25] Questionnaires were completed at screening, on day 1 of each 21-day cycle, at the end of treatment visit, 28 days after the last treatment, and at follow-up for disease progression. Scoring followed the EORTC QLQ-C30 scoring algorithm.[25] A linear transformation was used to standardize raw scores from 0 to 100.[25] A higher score on the GHS/QoL scale and the functioning scales indicated a higher ('better') level of functioning and HRQoL.[25]

Statistical Analyses

All analyses were descriptive and exploratory. *P* values were nominal and not adjusted for multiple testing. In line with the primary PFS and OS analysis of the trial, all models were adjusted for the following stratification factors: regorafenib pretreatment; time from onset of metastatic disease; and region. Five different methods were used to analyze the EORTC QLQ-C30 data.

Methods Using Group Means

These methods analyze the scales as continuous endpoints (longitudinal analyses), compute the mean of each treatment arm, and compare these means (Table 1). We used 2 of these methods to analyze data from LUME-Colon 1: the mixed-effects growth curve model and the MMRM.

Mean Treatment Difference for GHS/QoL and Functional Scale Scores Based on Mixed-effects Growth Curve Models. This method assessed changes in scores over time, with the average profile over time for each endpoint described by a piecewise linear regression model. For this particular study, weeks 3, 6, and 9 were selected to allow the slope to change; many patients were anticipated to drop out by week 12 owing to disease progression. All available data were used to fit the model.

Each model included 2 random effects: intercept and slope. To fit the model, the actual timepoints of the assessments were used. The mean score for each HRQoL endpoint was calculated from the area under the estimated growth curve up to the median follow-up time, and was interpreted as the average HRQoL score until the median follow-up time per group.

Mean Treatment Difference for GHS/QoL and Functional Scale Scores Based on MMRM. This longitudinal model also assessed changes in scores over time. It shows a pairwise comparison of mean scores between treatment arms for each timepoint/visit. The model included fixed categorical effects of treatment, stratification factors, visit, treatment-by-visit interaction, baseline, and baseline-by-visit interaction. The effects of visit and treatment-by-visit interaction allowed the model to incorporate time profiles without assuming a specific shape over time, in contrast to the piecewise linear shape in the mixed-effects growth curve model. The analysis was restricted to weeks 3, 6, 9, and 12. An unstructured covariance matrix was used to model the within-patient measurements.

Methods Assessing Clinically Relevant Changes at the Patient Level

These methods show differences between treatment arms based on the proportion of patients with an on-treatment improvement or a delay in worsening of the disease (Table 1).

Deteriorations Analyzed as TTD. In this evaluation, TTD was considered to be the time elapsed until a minimal clinically important deterioration (ie, of 10 points from baseline) occurred.[26] Cox regression, log-rank tests, and Kaplan-Meier curves were used as standard analysis methods. A stratified log-rank test, using the same stratification as for PFS, was used to investigate whether a delay in progression translated into a delayed worsening in HRQoL. Patients who died before their condition deteriorated were assumed to have deteriorated at the time of death. If a patient dropped out without experiencing deterioration or death, they were censored in this analysis. This analysis assumed that censoring was independent of deterioration in HRQoL.

Improvements Analyzed as Status Change (Improved, Stable, or Worsened). Status change assessed the proportion of patients who improved (even if this was temporary), stabilized, or worsened without any improvement for each EORTC QLQ-C30 scale.

Table 1 Summary of Methods That Can Be Used to Assess EORTC QLQ-C30

Methods	Description	MCID Required	One Event vs. Over Time
Methods using group means			
Mixed-effects growth curve models	• Mean score over time, typically assessed at selected timepoints	No	Over time
MMRM	• Estimates fixed effects: a visit effect, an arm effect, and a treatment-by-visit interaction. Random effects (individual trends) are also introduced[28] • This model accounts for the association of repeated measures made on the same patient[28]	No	Over time
Methods assessing clinically relevant changes at the patient level			
TTD	• Survival-based analyses[19] • Describes time from baseline until first instance of patient deterioration of at least one MCID unit[3] • Patients with no deterioration before they are lost to follow-up are censored at the time of the last follow-up or the last HRQoL assessment[3]	Yes	One event
TUDD	• Survival-based analysis[19] • Describes time from baseline until definitive deterioration (ie, after which point there is no improvement of $>$ 1 MCID unit compared with baseline value or if the patient dropped out after deterioration, resulting in missing data)[3]	Yes	One event
Status change	• Analysis of proportion of patients improving, worsening, or being stable, defined by change $\geq$ MCID • Comparison of percentage of patients being stable and/or improved vs. worsened	Yes	One event
Responder analysis	• Comparison of proportion of responders, defined as patients with mean improvement of $\geq$ 1 MCID during follow-up time of the study[30]	Yes	Over time
Time to improvement	• Time until first improvement $\geq$ MCID	Yes	One event

Abbreviations: EORTC QLQ-C30 = European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire; HRQoL = health-related quality of life; MCID = minimal clinically important difference; MMRM = mixed-effects model for repeated measurements; TTD = time to deterioration; TUDD = time until definitive deterioration.

Patients were categorized as having perceived improvement at least once (improvement of $\geq$ 10 units from baseline at any time during the study), regardless of potential worsening at other timepoints; worsened (decrease of $\geq$ 10 units and no improvement at any time during the study); or remained stable (neither showing improvement nor deterioration of $\geq$ 10 units at any time during the study). This endpoint was analyzed as a binary endpoint, comparing the proportion of patients with improvements between treatment groups using logistic regression.

Responder Analysis. In contrast to the status change analysis, the responder analysis was based on the average change from baseline over the entire observation period. A responder was defined as a patient who achieved an average $\geq$ 10-point increase in score from baseline over the duration of follow-up. The responder analysis compared the proportions of patients in each treatment arm, as was done for status change analysis.

Results

Patient Population and Baseline Characteristics

Patient demographics and baseline clinical characteristics have been reported previously and were mostly balanced between treatment groups.[22]

Completion of EORTC QLQ-C30 Questionnaires

During the first 12 cycles of treatment, EORTC QLQ-C30 questionnaires were completed by $\geq$ 87.5% of patients. In the nintedanib plus BSC arm and the placebo plus BSC arm, 82.7% and 80.4% of patients, respectively, completed questionnaires at the end of treatment.

Mean Treatment Difference for GHS/QoL and Functional Scale Scores

Baseline mean scores for GHS/QoL and physical functioning were comparable between treatment arms. Baseline mean scores for GHS/QoL were 64.9 $\pm$ 19.6 versus 65.3 $\pm$ 20.5 for the nintedanib and placebo arms, respectively. Baseline mean scores for physical functioning were 80.1 $\pm$ 18.4 in the nintedanib arm and 79.8 $\pm$ 17.9 in the placebo arm.

Mean treatment difference (ie, the difference in the area under the curve in the mixed-effects growth curve model, adjusted for stratification factors) favored nintedanib plus BSC over placebo plus BSC for physical functioning and social functioning. Numerical improvements were reported for GHS/QoL, role, and emotional and cognitive functioning (Figure 1).

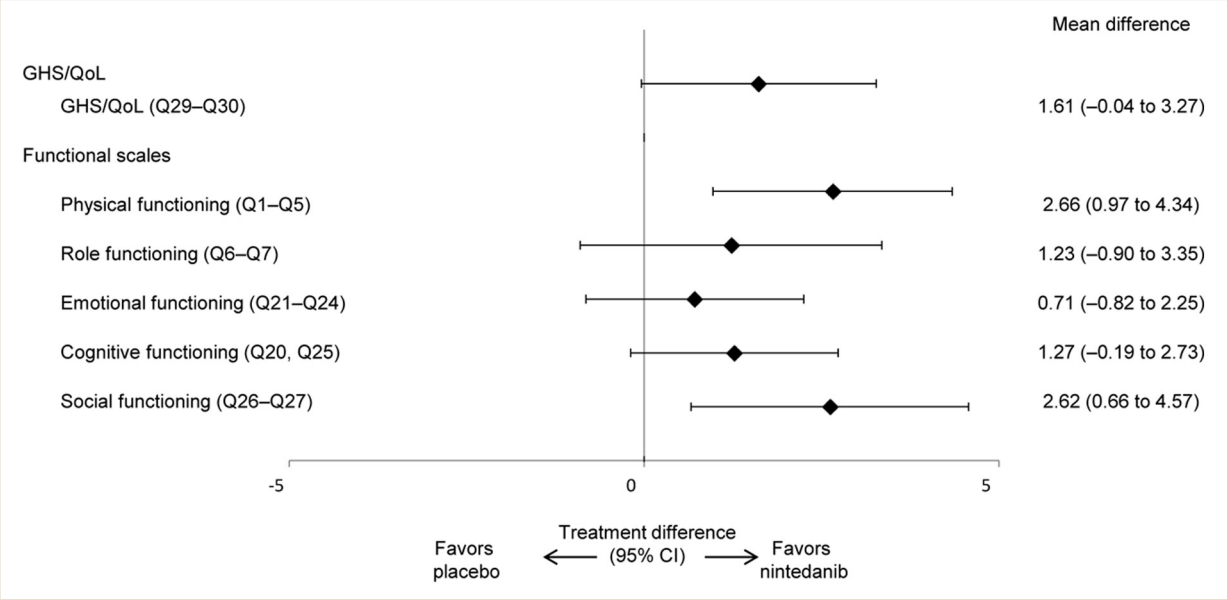
MMRM Analysis

Numerical improvements in the GHS/QoL scores and all functioning scores were reported in the nintedanib group versus the placebo group (Figures 2 and 3). Only physical functioning showed a significant between-group difference at all timepoints (Figure 3). Overall, a worsening in GHS/QoL and physical functioning was observed from weeks 3 to 12; this illustrates the expected, natural course of the disease (Figure 2).

There were significant between-group differences in physical functioning in favor of the nintedanib group at weeks 3, 6, 9, and 12 (Figure 3B).Treatment differences were similar between weeks 3 to 6 and 9 to 12, respectively, with larger differences in magnitude reported later in therapy (ie, at weeks 9 and 12) (Figure 3B).

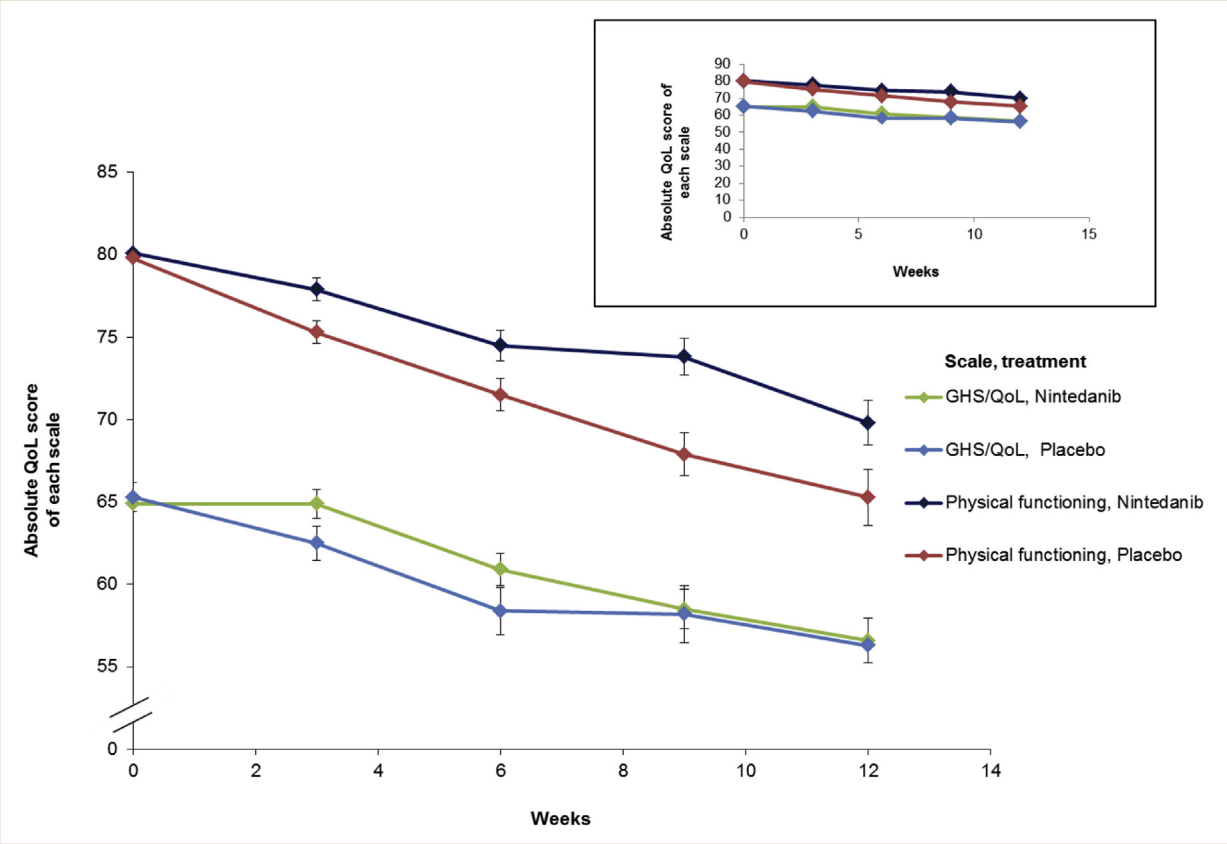
Between-treatment group differences in emotional functioning scores followed a similar trend to those observed for physical

Figure 1 Longitudinal Analysis: Mean Treatment Difference for GHS/QoL and Functional Scale Scores



Abbreviations: CI = confidence interval; GHS = global health status; Q = question; QoL = quality of life.

Figure 2 MMRM Analysis: Adjusted Mean Score of GHS/QoL and Physical Functioning. Figure Without Break in Axis Is Shown Inset. Baseline Values Are From Longitudinal Analysis and Are Shown for Reference. Error Bars Are ± SE



Abbreviations: GHS = global health status; MMRM = mixed-effects model for repeated measurements; QoL = quality of life.

functioning, although between-group differences were only significant at week 9 (see Supplemental Table 3 in the online version) (Figure 3D). Between-group differences in cognitive functioning and social functioning are shown in Figures 3E and F, respectively. In contrast to other endpoints, there were no significant between-treatment group differences in role functioning at any of the timepoints (see Supplemental Table 3 in the online version) (Figure 3C).

TTD

There was a general trend in favor of the nintedanib group compared with the placebo group for all functional scales apart from role functioning; there was also a trend in favor of the nintedanib group for GHS/QoL (Figures 4 and 5). Cognitive functioning showed a significant difference between treatment arms (Figure 5).

Status Change (Proportion of Patients Improved vs. Not Improved [Stable or Worsened])

The proportion of patients with improved GHS/QoL and physical functioning was significantly higher in the nintedanib group than in the placebo group (Figure 6), with an odds ratio equating to an improvement of 56% and 54%, respectively. There was a general trend in favor of nintedanib versus placebo for all other functional scales, with cognitive and social functioning also showing a significant difference between treatment groups (Figure 6).

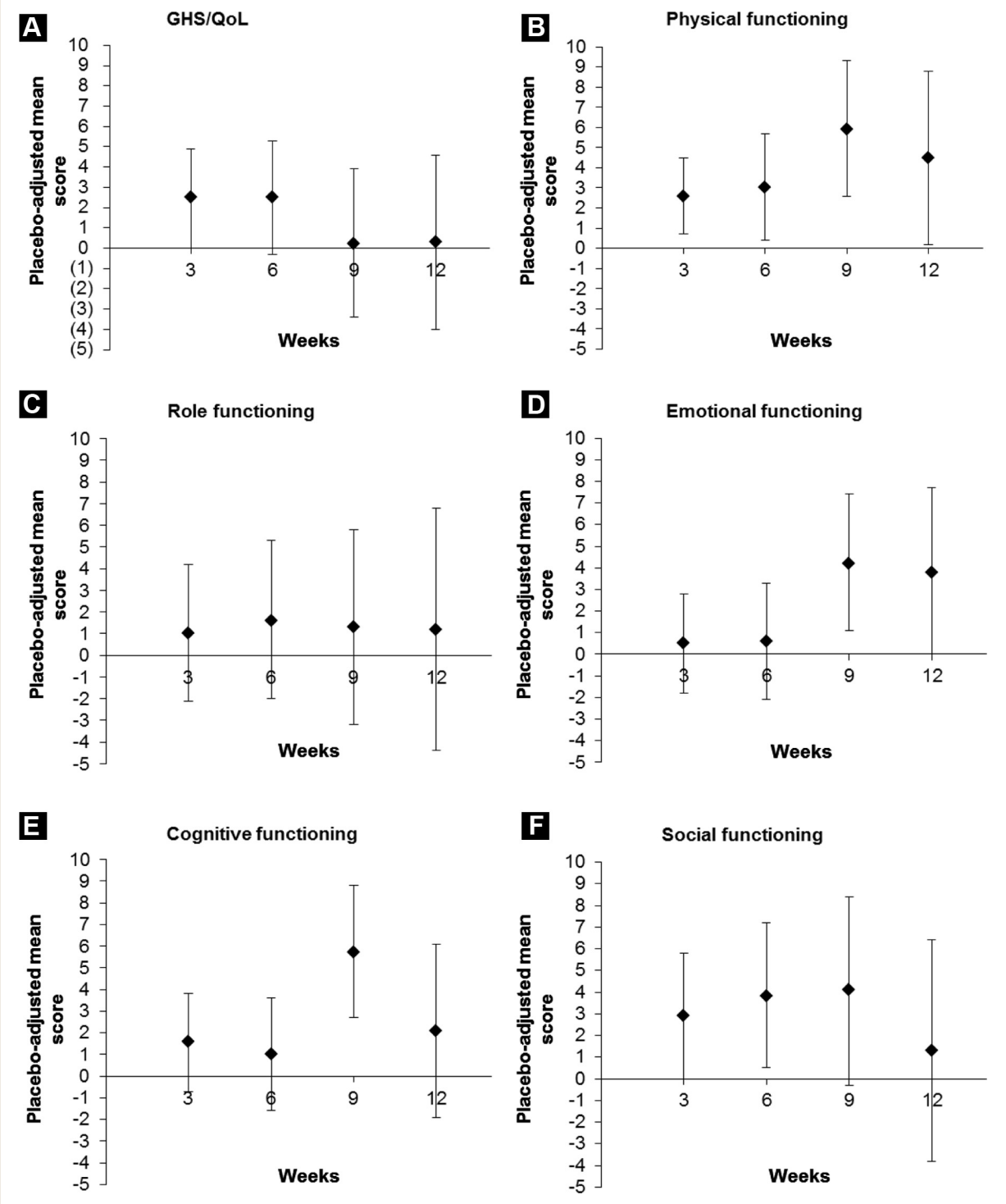
Responder Analysis

On average over the treatment period, only a small proportion of patients perceived an improvement in HRQoL and functioning scores (5.3%-13.8% in the placebo arm and 6.9%-17.2% in the nintedanib arm) (see Supplemental Table 4 in the online version). Nevertheless, the nintedanib group had a somewhat higher response rate in GHS/QoL and physical functioning scores than the placebo group, with an odds ratio equating to an improvement of 46% and 32%, respectively (see Supplemental Table 4 in the online version). There were higher responder rates with nintedanib versus placebo in the cognitive and social functioning scales than in other functional scales, equating to a relative improvement of 54% and 29%, respectively (see Supplemental Table 4 in the online version). However, the confidence intervals included 1 for all scores.

Discussion

The LUME-Colon 1 study failed to meet both co-primary endpoints: there was no survival benefit and a significant but modest increase in PFS with nintedanib versus placebo. PRO data

Figure 3 MMRM Analysis: Mean Treatment Difference. Plots Shown Are Placebo-corrected Means With Upper and Lower Confidence Intervals

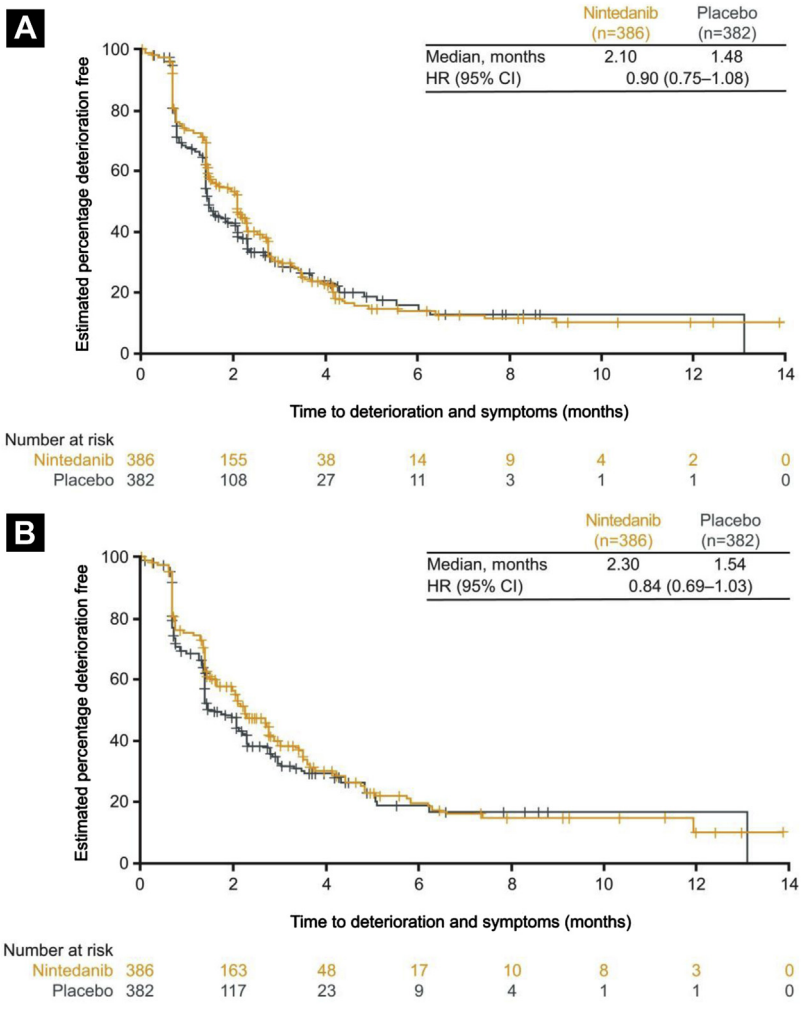


Abbreviations: GHS = global health status; MMRM = mixed-effects model for repeated measurements; QoL = quality of life.

from the study, reported here, indicate that patients' overall GHS/QoL status and physical functioning were not impaired, and were improved in some respects, by treatment with nintedanib compared with placebo.

Overall, results from the different methods of analysis showed an apparent benefit of treatment with nintedanib versus placebo in various PROs, in particular in physical functioning. The longitudinal analysis using the mixed-effects growth curve model included

Figure 4 Kaplan-Meier Curves of Time to Deterioration of GHS/QoL Score (A) and Physical Functioning Score (B)



Abbreviations: CI = confidence interval; GHS = global health status; HR = hazard ratio; QoL = quality of life.

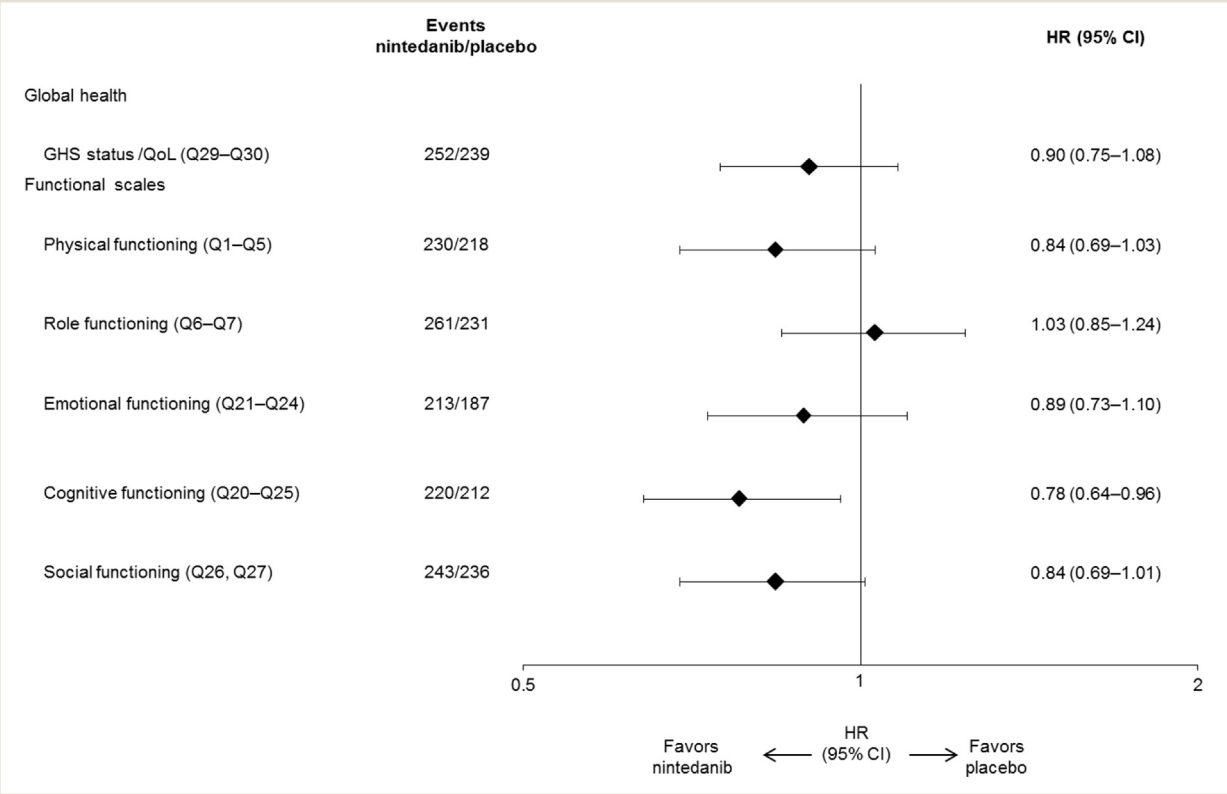
the first 9 weeks of treatment. Results showed that average physical functioning was better with nintedanib than with placebo during this time. MMRM confirmed this pattern through weeks 3, 6, 9, and 12, with larger improvements at weeks 9 and 12. The 'status change' analysis, which focused on temporary but clinically relevant changes in individual patients at specific timepoints during therapy, supported the outputs of the longitudinal analysis. It showed that there were statistically significant and clinically meaningful improvements in physical functioning, GHS/QoL, and cognitive and social functioning with nintedanib versus placebo.

In contrast to the status change analysis, the responder analysis had a stronger criterion of a mean improvement of $\geq$ 10 points averaged over all timepoints in the study. This analysis did not show a statistically significant difference between treatment groups, perhaps reflecting the short time to disease progression in this study population. TTD in physical functioning was numerically longer with nintedanib than with placebo, although there was no significant difference between treatment arms. Data from the individual items of the physical functioning assessment indicate that patients experienced delayed deterioration in their ability to undertake strenuous activity (eg, heavy lifting) and take a walk outside, and in their need for help when eating and dressing (see Supplemental Figure 3 in the online version). Looking at individual items in this way can contribute to our understanding of patients' experience while on treatment.

The variety of statistical methods used to analyze data obtained using EORTC QLQ-C30 can make it difficult to interpret the likely impact of those data in the clinic.[27] This lack of standards for analyzing PRO and HRQoL data has been recognized, and a consortium initiative (the Setting International Standards in Analysing Patient-Reported Outcomes and Quality of Life Endpoints Data) has been established. This consortium, directed by the EORTC, aims to establish recommendations on how to standardize the analysis of HRQoL and other PRO data in randomized cancer drug trials.[27]

The outputs of the different statistical methods used to analyze EORTC QLQ-C30 in this study demonstrate that there are

Figure 5 TTD: GHS/QoL and Functioning Scales. Results Presented on a Logarithmic Scale to Base 4



Abbreviations: CI = confidence interval; GHS = global health status; HR = hazard ratio; Q = question; QoL = quality of life; TTD = time to deterioration.

different ways to evaluate what patients experience in everyday life while on treatment. Interpretation of these different analyses requires context: in this study, it is important to note the rapid deterioration in QoL and physical functioning experienced by patients in both treatment arms, and the short PFS. Clinical interpretation of the results of these methods requires us to bear in mind the strengths and limitations of each one, as follows.

Mixed-effects Growth Curve and MMRM Models

These analyses offer an overall comparison of the treatment arms, taking the longitudinal course of the disease and treatment into account. Both models assume that data are missing at random. This assumption is violated if patients who drop out have worse HRQoL scores than those who continue in the study, which is a common scenario. However, a key advantage of these methods is that group means can be analyzed without a defined MCID. Furthermore, all the information from the questionnaire can be analyzed directly, without the need to classify the collected data as an improvement or a deterioration. In addition, mixed-effects growth curve models are able to incorporate the actual timepoint of a measurement, even if there are many assessments during long-term follow-up, whereas the MMRM uses the planned timepoints of the assessments. MMRM is regarded as a well-established and robust approach, particularly if there are missing data (as long as those data are missing at random).[28]

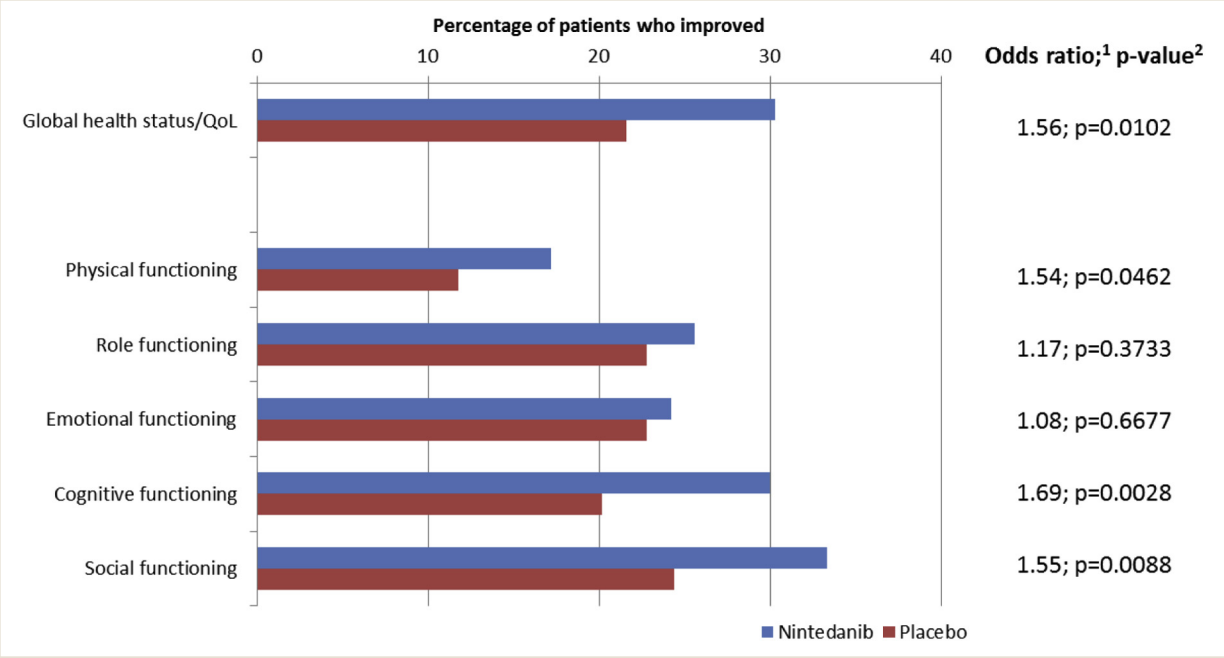
One disadvantage of mixed-effects growth curve and MMRM models is that the outputs can be difficult to interpret: clinically relevant criteria for comparing group means are needed, which are often more difficult to substantiate than a MCID on the individual patient level. Another disadvantage is that during long-term follow-up, the assignment of timepoints to visits might be challenging if clinic visits are delayed.

TTD

Among the strengths of TTD is its use of MCID. For well-established questionnaires such as the EORTC QLQ-C30, widely accepted MCIDs are available. This makes TTD a clinically relevant measure that is easy to interpret and apply. TTD is a particularly meaningful analysis if a deterioration in HRQoL is expected long term and the therapy tested is expected to delay progression and deterioration of HRQoL. Analysis of TTD mirrors the analysis of PFS, complementing the interpretation of this common endpoint in oncology trials. Finally, TTD analyses can be adjusted using censoring to take into account missing data, patient deaths, and changes in patients' individual perception of their QoL over time.[28]

One of the disadvantages of TTD analyses is the need to clearly define what constitutes deterioration.[28] Another point is that the analysis may be less robust if patients drop out because of progressive disease but do not experience deterioration according to the MCID; this would mean that censoring could introduce bias into the analysis. Furthermore, when interpreting the results of a TTD analysis, it is necessary to keep in mind that only the time to first deterioration is taken into account – information on long-term

Figure 6 Status Change: Proportion of Patients With Improvement in Functional Scales and GHS/QoL Scores. [1]Odds Ratio >1 Favors Nintedanib; [2]*P* Value Corresponds to Patients Who Improved vs. Patients Who Did Not Improve (ie, Stable Plus Worsened Categories Combined)



Abbreviations: GHS = global health status; QoL = quality of life.

treatment effects after the first deterioration is not considered. Hence, TTD analysis might be most appropriate if the first deterioration is of interest, or if the natural course of disease suggests a continuous deterioration. Finally, TTD analyses assume that censoring is non-informative; if data are not missing at random before a deterioration occurs, this assumption is violated.

Status Change

The main advantage of the status change method is its simplicity: it looks at clinically meaningful changes in HRQoL at the level of each individual patient. Depending on how status change is defined, even a single measured change from baseline can be sufficient to classify a patient as 'improved.' In severe, incurable diseases, the overall course of the disease is one of deterioration, and so a status change analysis can reveal at least temporary improvements that may be experienced by some patients.

A potential disadvantage of this method is that it is susceptible to bias in favor of the treatment arm that had a longer follow-up. Status change analyses are therefore most appropriate in studies that have similar expected observation periods in all treatment groups.

Responder Analysis

Responder analyses require a mean improvement over all timepoints of at least the MCID. It is more stringent than the status change analysis, and does not just take into account temporary improvements. It is appropriate if a mean improvement can be expected over the entire course of study treatment.

Disadvantages of responder analyses include that the definition of response needs to be based on a validated MCID.[29] In addition, this type of analysis has reduced power compared with an analysis based on continuous data; also, it is more challenging to achieve a response with longer follow-up if there is a fast deterioration in HRQoL owing to the underlying disease.

General Points on the Interpretation of Different Analyses

In some cases, the interpretation of methods that are based on clinically relevant changes can be misleading. For example, patients who experience a different QoL profile over time could have the same TTD if they reach the MCID at the same timepoint (see Supplemental Figure 1 in the online version). Alternatively, a patient whose disease status deteriorates rapidly but then improves could have a shorter TTD than a patient whose QoL remains stable for some time but then deteriorates (see Supplemental Figure 2 in the online version). Similarly, a patient with a short improvement in QoL who drops out because of disease progression and then has a rapid deterioration in HRQoL after the end of follow-up could be categorized as a 'responder.' The potential for misinterpretation highlights the importance of analyzing HRQoL data using different, complementary methods to obtain an accurate picture of patients' experiences.

Analysis of QoL endpoints needs to be meaningful in context, which includes consideration of the type and severity of symptoms and the course of treatment and disease. In this study, the patient population had late-stage CRC, with a high symptom burden, fast deterioration of QoL, and short survival. Delay of progression and deterioration while maintaining HRQoL and functioning is the

primary goal of therapy. Given the considerations listed in this Discussion, analysis of TTD, complemented by a longitudinal analysis and an analysis of status change, may be most useful for characterizing patients' experience on study treatment.

Conclusions

PRO data from LUME-Colon 1 showed that the overall GHS/QoL status was not impaired by active treatment with nintedanib, and that patients perceived some HRQoL benefits with nintedanib compared with placebo. This is in line with the modest PFS benefit observed in the study. Careful consideration and comparison of different methods of analysis of EORTC QLQ C-30 are required in settings such as metastatic refractory CRC in order to support clinical interpretation and application of findings.

Clinical Practice Points

- PROs are an essential part of clinical trials, and in particular play a key role in interpreting the benefit of a novel treatment from the patient's perspective.
- We used EORTC QLQ-C30 data from the LUME-Colon 1 study to compare and contrast the different statistical methods that can be used to evaluate these types of data. The different analyses complement each other.
- These analyses show that, in LUME-Colon 1, overall GHS/QoL status was not impaired by active treatment with nintedanib compared with placebo. In fact, patients even perceived some benefits with nintedanib compared with placebo, in particular with regard to physical functioning.
- However, there was no survival benefit and only a modest increase in PFS with nintedanib versus placebo in this study.
- Careful consideration and comparison of different methods of analysis of EORTC QLQ C-30 are required in settings such as metastatic refractory CRC in order to support clinical interpretation and application of findings.

Acknowledgments

This work was supported by Boehringer Ingelheim. Medical writing assistance, supported financially by Boehringer Ingelheim, was provided by Syneos Health Communications during the preparation of this manuscript.

Disclosure

H.-J.L. has had an advisory role for Bayer, Merck Serono, Pfizer, and Roche, and has received honoraria from Bayer, Boehringer Ingelheim, Merck Serono, and Roche. G.A. has received honoraria for advisory roles, travel grants and research grants (past five years) from Hoffman La-Roche, Bristol Myers Squibb, Bayer, Servier, Amgen, Merck Serono and Menarini; and his institution has received honoraria due to his investigator contribution in clinical trials from Bayer, Servier, Novartis, Boehringer Ingelheim, Boston Pharmaceuticals, Hoffman La-Roche and Genentech. T.Y.'s institution has received research funds from Boehringer Ingelheim and GlaxoSmithKline. S.L. has received research funds from Sanofi; had an advisory role with Bayer and Amgen; and attended a speakers' bureau for Roche and Lilly. A.F. has received research funds paid to his institute from Amgen, Bayer, Merck, Roche, Sanofi, MSD, and Servier; has had an advisory role for Amgen, Bayer, Bristol-Myers Squibb, Lilly, Merck, Roche, Sanofi, and Servier; and received honoraria from Celgene, Lilly, Merck, Roche, and Servier. A.S. has had an advisory role for Amgen, Bayer, Celgene, Roche, Merck Serono, Sanofi, and Servier, and has attended a speakers' bureau for Amgen, AstraZeneca, Bayer, Bristol-Myers Squibb, Celgene, Lilly, Merck Serono, Roche, Sanofi, and Takeda. E.V.C.'s institution has received research funds from Amgen, Bayer, Boehringer Ingelheim, Celgene, Ipsen, Lilly, Merck, Merck KGaA, Novartis, Roche, Sanofi, and Servier and has had an advisory role for Bayer, Lilly, Roche, and Servier. C.H., B.P., F.V., and I.G. are employees of Boehringer Ingelheim. M.L.L. has nothing to declare.

Supplemental Data

Supplemental tables and figures accompanying this article can be found in the online version at https://doi.org/10.1016/j.clcc.2019.08.005.

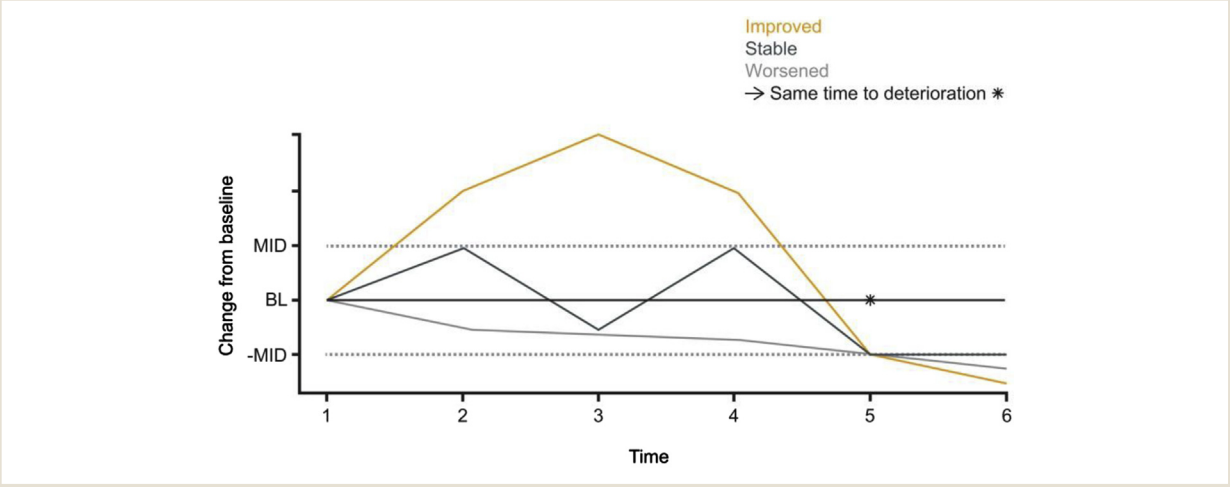
References

1. Committee for Medicinal Products for Human Use, European Medicines Agency. Reflection paper on the use of patient reported outcome (PRO) measures in oncology studies. Draft. June 17, 2014, Available at: www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2014/06/WC500168852.pdf. Accessed: August 31, 2017.
2. Brettschneider C, Luhmann D, Raspe H. Informative value of patient reported outcomes (PRO) in health technology assessment (HTA). *GMS Health Technol Assess* 2011; 7:Doc01.
3. Anota A, Hamidou Z, Paget-Bailly S, et al. Time to health-related quality of life score deterioration as a modality of longitudinal analysis for health-related quality of life studies in oncology: do we need RECIST for quality of life to achieve standardization? *Qual Life Res* 2015; 24:5-18.
4. Bell ML, Kenward MG, Fairclough DL, Horton NJ. Differential dropout and bias in randomised controlled trials: when it matters and when it may not. *BMJ* 2013; 346:e8668.
5. Pe M, Dorme L, Coens C, et al, Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data Consortium (SISAQOL). Statistical analysis of patient-reported outcome data in randomised controlled trials of locally advanced and metastatic breast cancer: a systematic review. *Lancet Oncol* 2018; 19:e459-69.
6. Giesinger JM, Kieffer JM, Fayers PM, et al, EORTC Quality of Life Group. 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.
7. Yamaguchi K, Ando M, Ooki A, et al. Quality of life analysis in patients with RAS wild-type metastatic colorectal cancer treated with first-line cetuximab plus chemotherapy. *Clin Colorectal Cancer* 2017; 16:e29-37.
8. Pinto C, Di Fabio F, Rosati G, et al. Observational study on quality of life, safety, and effectiveness of first-line cetuximab plus chemotherapy in KRAS wild-type metastatic colorectal cancer patients: the ObservEr Study. *Cancer Med* 2016; 5:3272-81.
9. Quidde J, Hegewisch-Becker S, Graeven U, et al. Quality of life assessment in patients with metastatic colorectal cancer receiving maintenance therapy after first-line induction treatment: a preplanned analysis of the phase III AIO KRK 0207 trial. *Ann Oncol* 2016; 27:2203-10.
10. Hamidou Z, Chibaudel B, Hebbar M, et al. Time to definitive health-related quality of life score deterioration in patients with resectable metastatic colorectal cancer treated with FOLFOX4 versus sequential dose-dense FOLFOX7 followed by FOLFIRI: the MIROX randomized phase III trial. *PLoS One* 2016; 11: e0157067.
11. Yoshino T, Komatsu Y, Yamada Y, et al. Randomized phase III trial of regorafenib in metastatic colorectal cancer: analysis of the CORRECT Japanese and non-Japanese subpopulations. *Invest New Drugs* 2015; 33:740-50.
12. Qin S, Kim TW, Yau TC, et al. Effects of regorafenib therapy on health-related quality of life (HRQoL) in patients with metastatic colorectal cancer (mCRC) in the phase III CONCUR trial. *J Clin Oncol* 2015; 33(3 Suppl):697.
13. Thaler J, Karthaus M, Mineur L, et al. Skin toxicity and quality of life in patients with metastatic colorectal cancer during first-line panitumumab plus FOLFIRI treatment in a single-arm phase II study. *BMC Cancer* 2012; 12:438.
14. Au HJ, Karapetis CS, O'Callaghan CJ, et al. Health-related quality of life in patients with advanced colorectal cancer treated with cetuximab: overall and KRAS-specific results of the NCIC CTG and AGITG CO.17 Trial. *J Clin Oncol* 2009; 27:1822-8.
15. Tol J, Koopman M, Cats A, et al. Chemotherapy, bevacizumab, and cetuximab in metastatic colorectal cancer. *N Engl J Med* 2009; 360:563-72.
16. Sobrero AF, Maurel J, Fehrenbacher L, et al. EPIC: phase III trial of cetuximab plus irinotecan after fluoropyrimidine and oxaliplatin failure in patients with metastatic colorectal cancer. *J Clin Oncol* 2008; 26:2311-9.

17. Rosati G, Cordio S, Tucci A, et al. Phase II trial of oxaliplatin and tegafur/uracil and oral folinic acid for advanced or metastatic colorectal cancer in elderly patients. *Oncology* 2005; 69:122-9.
18. Grothey A, Van Cutsem E, Sobrero A, et al, CORRECT Study Group. Regorafenib monotherapy for previously treated metastatic colorectal cancer (CORRECT): an international, multicentre, randomised, placebo-controlled, phase 3 trial. *Lancet* 2013; 381:303-12.
19. Williams P, Fofana F, de la Loge C, et al. *Methods for Analyzing Patient-reported Outcomes in Oncology Trials. Poster presented at ISPOR* 2016, Vienna, Austria. Poster PRM233.
20. Hilberg F, Roth GJ, Krssak M, et al. BIBF 1120: triple angiokinase inhibitor with sustained receptor blockade and good antitumor efficacy. *Cancer Res* 2008; 68:4774-82.
21. *Boehringer Ingelheim. Data on File* 2018.
22. Van Cutsem E, Yoshino T, Lenz HJ, et al. Nintedanib for the treatment of patients with refractory metastatic colorectal cancer (LUME-Colon 1): a phase III, international, randomized, placebo-controlled study. *Ann Oncol* 2018; 29:1955-63.
23. Mross K, Buchert M, Frost A, et al. Vascular effects, efficacy and safety of nintedanib in patients with advanced, refractory colorectal cancer: a prospective phase I subanalysis. *BMC Cancer* 2014; 14:510.
24. Van Cutsem E, Prenen H, D'Haens G, et al. A phase I/II, open-label, randomised study of nintedanib plus mFOLFOX6 versus bevacizumab plus mFOLFOX6 in first-line metastatic colorectal cancer patients. *Ann Oncol* 2015; 26:2085-91.
25. Fayers P, Aaronson N, Bjordal K, et al. *The EORTC QLQ-C30 Scoring Manual.* 3rd ed. Brussels: European Organisation for Research and Treatment of Cancer; 2001.
26. Osoba D, Bezjak A, Brundage M, Pater J, National Cancer Institute of Canada Clinical Trials Group. Evaluating health-related quality of life in cancer clinical trials: the National Cancer Institute of Canada Clinical Trials Group experience. *Value Health* 2007; 10(Suppl 2):S138-45.
27. Bottomley A, Pe M, Sloan J, et al, Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data (SISAQOL) consortium. Analysing data from patient-reported outcome and quality of life endpoints for cancer clinical trials: a start in setting international standards. *Lancet Oncol* 2016; 17:e510-4.
28. Bonnetain F, Fiteni F, Efficace F, Anota A. Statistical challenges in the analysis of health-related quality of life in cancer clinical trials. *J Clin Oncol* 2016; 34:1953-6.
29. Snapinn SM, Jiang Q. Responder analyses and the assessment of a clinically relevant treatment effect. *Trials* 2007; 8:31.
30. Lin Y. Robust inference for responder analysis: innovative clinical trial design using a minimum p-value approach. *Contemp Clin Trials Commun* 2016; 3:65-9.
31. Lang I, Kohne CH, Folprecht G, et al. Quality of life analysis in patients with KRAS wild-type metastatic colorectal cancer treated first-line with cetuximab plus irinotecan, fluorouracil and leucovorin. *Eur J Cancer* 2013; 49:439-48.

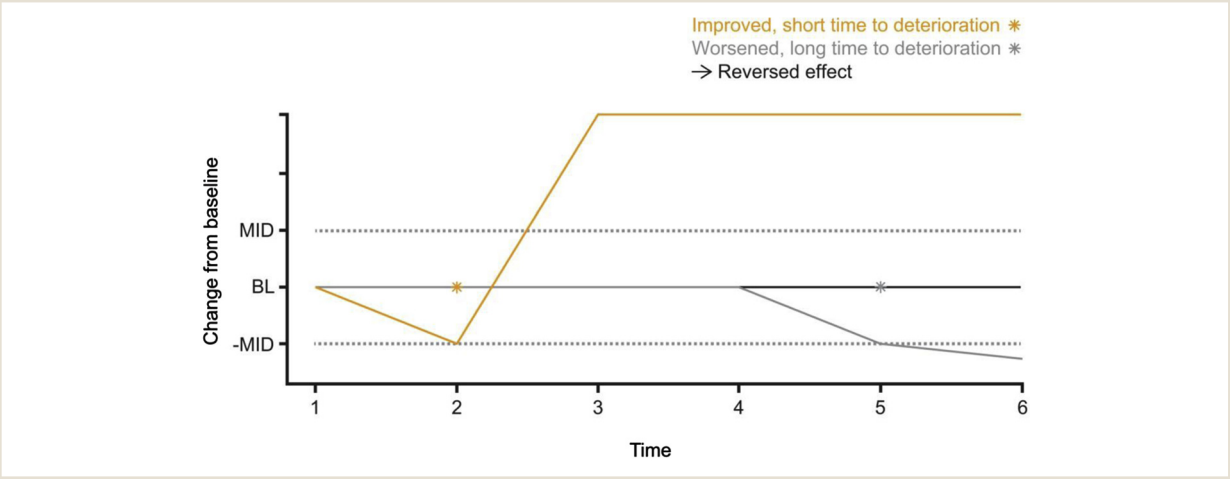
Supplemental Data

Supplemental Figure 1 Individual Patient Profiles Over Time With Same TTD



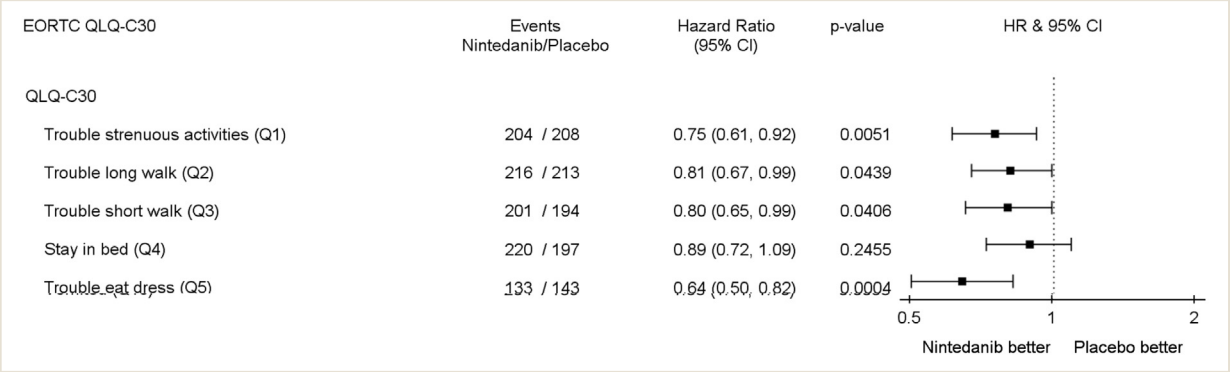
Abbreviations: BL = baseline; MID = minimal important difference; TTD = time to deterioration.

Supplemental Figure 2 Individual Patient Profiles With Short and Long TTD



Abbreviations: BL = baseline; MID = minimal important difference; TTD = time to deterioration.

Supplemental Figure 3 TTD: Physical Functioning Scale — Individual Items



Abbreviations: CI = confidence interval; EORTC QLQ-C30 = European Organization for Research and Treatment of Cancer Quality of Life Questionnaire; HR = hazard ratio; Q = question; TTD = time to deterioration.

Supplemental Table 1 Analysis of the EORTC QLQ-C30 in CRC Trials of Systemic Drug Therapy

Study	Difference Between Arms at Specified Timepoints	Change From Baseline to Specified Timepoint	TTD	TUDD	Longitudinal Analysis	Responder	Status Change
Yamaguchi et al, 2017[7] and Lang et al, 2013[31] Cetuximab plus chemotherapy: CRYSTAL study	×	×			×		×
Pinto et al, 2016[8] Cetuximab plus chemotherapy: ObservEr Study		×			×		
Quidde et al, 2016[9] Fluoropyrimidine plus bevacizumab: AIO KRK 0207 trial	×				×	×	×
Hamidou et al, 2016[10] FOLFOX4 versus sequential dose-dense FOLFOX7 followed by FOLFIRI: MIROX trial				×			
Grothey et al, 2013[18] and Yoshino et al, 2015[11] Regorafenib: CORRECT trial		×			×		
Qin et al, 2015[12] Regorafenib: CONCUR trial		×			×		
Thaler et al, 2012[13] Panitumumab plus FOLFIRI		×			×		
Au et al, 2009[14] Cetuximab: NCIC CTG and AGITG CO.17 trial		×	×			×	×
Tol et al, 2009[15] Cetuximab plus chemotherapy and bevacizumab		×					
Sobrero et al, 2008[16] Cetuximab plus irinotecan: EPIC trial		×			×		
Rosati et al, 2005[17] Oxaliplatin plus tegafur/uracil and oral folinic acid		×					

Abbreviations: CRC = colorectal cancer; EORTC QLQ-C30 = European Organization for Research and Treatment of Cancer Quality of Life Questionnaire; TTD = time to deterioration; TUDD = time until definitive deterioration.

Supplemental Table 2 Predefined HRQoL Measures of Interest in LUME-Colon 1

Scale	Item/Question
GHS/QoL[a]	1 How would you rate your overall health during the past week? 2 How would you rate your overall QoL during the past week?
Physical functioning[b]	1 Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase? 2 Do you have any trouble taking a long walk? 3 Do you have any trouble taking a short walk outside of the house? 4 Do you need to stay in bed or on a chair during the day? 5 Do you need help with eating, dressing, washing yourself, or using the toilet?

Abbreviations: GHS = global health status; HRQoL = health-related quality of life; QoL = quality of life.
[a]For both questions, the patient can circle a number between 1 (very poor) and 7 (excellent).
[b]There are 4 response categories: 1 = not at all, 2 = a little, 3 = quite a bit, and 4 = very much. Questionnaires were completed at timepoints specified in the protocol: at screening, day 1 of each 21-day cycle, at the end of treatment visit, at the end of the residual period, and at follow-up for disease progression.

Supplemental Table 3 MMRM Analysis

	Difference From Placebo	
Endpoint	Adjusted Mean (SE)	95% CI
GHS/QoL (Q29, Q30)		
Week 3	2.5 (1.3)	0.0 to 4.9
Week 6	2.5 (1.4)	−0.3 to 5.3
Week 9	0.2 (1.9)	−3.4 to 3.9
Week 12	0.3 (2.2)	−4.0 to 4.6
Physical functioning (Q1-Q5)		
Week 3	2.6 (1.0)	0.7 to 4.5
Week 6	3.0 (1.4)	0.4 to 5.7
Week 9	5.9 (1.7)	2.6 to 9.3
Week 12	4.5 (2.2)	0.2 to 8.8
Role functioning (Q6-Q7)		
Week 3	1.0 (1.6)	−2.1 to 4.2
Week 6	1.6 (1.9)	−2.0 to 5.3
Week 9	1.3 (2.3)	−3.2 to 5.8
Week 12	1.2 (2.9)	−4.4 to 6.8
Emotional functioning (Q21-Q24)		
Week 3	0.5 (1.2)	−1.8 to 2.8
Week 6	0.6 (1.4)	−2.1 to 3.3
Week 9	4.2 (1.6)	1.1 to 7.4
Week 12	3.8 (2.0)	0.0 to 7.7
Cognitive functioning (Q20, Q25)		
Week 3	1.6 (1.1)	−0.7 to 3.8
Week 6	1.0 (1.3)	−1.6 to 3.6
Week 9	5.7 (1.6)	2.7 to 8.8
Week 12	2.1 (2.0)	−1.9 to 6.1
Social functioning (Q26, Q27)		
Week 3	2.9 (1.5)	0.0 to 5.8
Week 6	3.8 (1.7)	0.5 to 7.2
Week 9	4.1 (2.2)	−0.3 to 8.4
Week 12	1.3 (2.6)	−3.8 to 6.4

Results are from an MMRM model. Fixed effects include treatment, stratification factors, visit, treatment-by-visit interaction, baseline, and baseline-by-visit interaction. Number of patients contributing to the MMRM model in each treatment group: placebo, n = 356; nintedanib, n = 360.
Abbreviations: CI = confidence interval; GHS = global health status; MMRM = mixed-effects model for repeated measurements; Q = question; QoL = quality of life; SE = standard error.

Supplemental Table 4 Responder Analysis

Endpoint	Number of Responders (%)		Nintedanib vs. Placebo Odds Ratio (95% CI)
	Placebo (N[a] = 356)	Nintedanib (N[a] = 360)	
GHS/QoL			
GHS/QoL (Q29, Q30)	35 (9.8)	50 (13.9)	1.46 (0.92-2.31)
Functional scales			
Physical functioning (Q1-Q5)	19 (5.3)	25 (6.9)	1.32 (0.71-2.45)
Role functioning (Q6, Q7)	39 (11.0)	40 (11.1)	1.01 (0.63-1.61)
Emotional functioning (Q21-Q24)	45 (12.6)	46 (12.8)	1.00 (0.64-1.55)
Cognitive functioning (Q20, Q25)	31 (8.7)	46 (12.8)	1.54 (0.95-2.49)
Social functioning (Q26, Q27)	49 (13.8)	62 (17.2)	1.29 (0.86-1.94)

Results are from fitting a generalized linear regression model adjusted for treatment and stratifications factors used at randomization and a logit link (odds ratio). An odds ratio > 1 indicates a benefit to nintedanib.
Abbreviations: CI = confidence interval; GHS = global health status; Q = question; QoL = quality of life.
[a]Number of patients with baseline and at least one post-baseline assessment.