Psychometric Validation of the EORTC QLQ-HCC18 in Patients with Previously Treated Unresectable Hepatocellular Carcinoma

Daniel Serrano
Pharmerit North America LLC
Lauren Podger
Pharmerit North America LLC
Gisoo Barnes (Gisoo.barnes@beigene.com)
BeiGene
James Song
BeiGene
Boxiong Tang
BeiGene

Research Article

Keywords: Hepatocellular carcinoma, patient-reported outcomes, health-related quality of life, psychometric analysis, classical test theory, responsiveness, meaningful change, EORTC QLQ-HCC18

Posted Date: March 16th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-191917/v1

Version of Record: A version of this preprint was published at Quality of Life Research on September 13th, 2021. See the published version at https://doi.org/10.1007/s11136-021-02992-1.

Abstract

Purpose: Demonstrate the measurement properties of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Hepatocellular Carcinoma 18-question module (EORTC QLQ-HCC18) within a previously treated, unresectable HCC clinical trial population that was distinct from the published QLQ-HCC18 validation population.

Methods: Analyses were conducted using data from BGB-A317-208, an open label, international, clinical trial of the monoclonal antibody tislelizumab in adult HCC patients. The EORTC Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) and QLQ-HCC18 were assessed at baseline as well as weeks 3 and 9. Psychometric validation of the QLQ-HCC18 included reliability, construct validity, ability to detect change, and meaningful within-patient change (MWPC). Known-groups validity and MWPC analyses were also stratified on several pre-defined subpopulations.

Results: A total of 248 patients were included. The QLQ-HCC18 fatigue, nutrition, and index domains demonstrated acceptable internal consistency; acceptable test-retest reliability was found for fatigue, body image, nutrition, pain, sexual interest, and index domains. The QLQ-HCC18 fatigue domain achieved acceptable concurrent validity for 13 of 16 correlations; the index domain achieved acceptable concurrent validity for 15 of 16 correlations. Clear differentiation of the QLQ-HCC18 change scores between improvement and maintenance anchor groups were observed for body image, fatigue, pain, and index domains. Differentiation between deterioration and maintenance anchor groups were observed for fever and fatigue domains. MWPC point estimates defining improvement for the QLQ-HCC18 fatigue and index domains were -7.18 and -4.07, respectively; MWPC point estimates defining deterioration were 5.34 and 3.16, respectively.

Conclusions: The EORTC QLQ-HCC18 fatigue and index domains consistently demonstrated robust psychometric properties, supporting the use of these domains as suitable patient-reported endpoints within a previously treated, unresectable HCC patient population.

1.0 Introduction

Hepatocellular carcinoma (HCC) is a substantial global health challenge that accounts for 85–90% of all reported cases of liver cancer and is the fourth most common cause of cancer-related death [1]. In addition, between 80% and 90% of people worldwide with HCC have comorbid hepatitis B virus (HBV) and/or hepatitis C virus (HCV) infection [2, 3]. Most HCC cases (> 80%) occur in Eastern Asia and sub-Saharan Africa, with typical incidence rates of > 20 per 100,000 individuals; China alone accounts for approximately 50% of both new HCC cases and HCC-related deaths worldwide [4, 5]. Southern European countries, such as Spain, Italy, and Greece, have higher incidence rates (10 to 20 per 100,000 individuals) in comparison to Northern Europe and the Americas [4, 5].

Patients with unresectable HCC represent a population with great unmet medical need, having a 5-year overall survival (OS) rate of 18% [6]. These patients often report symptoms (i.e., muscle cramps, pain, fatigue, sleep dysfunction) severe enough to affect their health-related quality of life (HRQoL) [7]. Furthermore, these symptoms affecting HRQoL have been found to correlate with shorter OS [7–10]. As a result, there has been a shift towards increased recognition for the need to assess HRQoL alongside traditional clinical outcomes in HCC trials [11]. Several different questionnaires have been employed to measure HRQoL in studies of HCC [7]; however, only the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Hepatocellular Carcinoma 18-question module (EORTC QLQ-HCC18) was developed specifically to assess symptom burden and impact on HRQoL in people with HCC [12–14]. As it stands, there are limited published data demonstrating the

measurement properties of the QLQ-HCC18 within an unresectable HCC population, as well as within specific subpopulations including viral hepatitis comorbidities (comorbid HBV and HCV versus no comorbidity), line of therapy (second- versus third-line or greater), and geographic region (Asia versus Europe).

The objective of the current project was to validate the QLQ-HCC18 within the BGB-A317-208 trial population. Patients participating in BGB-A317-208 were treated with tislelizumab, a humanized immunoglobulin G4 (IgG)-variant monoclonal antibody affecting the immune checkpoint-inhibitory receptor known as programmed cell death protein-1 (PD-1) [15, 16]. The BGB-A317-208 context of use was distinct from the published QLQ-HCC18 validation population, with the former consisting of previously treated, unresectable HCC patients. Thus, following the US Food and Drug Administration (FDA) guidance [17, 18], analyses of the QLQ-HCC18 were conducted to evaluate psychometric precision (reliability, construct validity, ability to detect change, and meaningful within-patient change [MWPC]) within this patient population.

2.0 Methods

This validation study was conducted using BGB-A317-208 trial data. BGB-A317-208 (NCT0341989) was an open label, multicenter, international, Phase 2 clinical trial assessing efficacy and safety of the anti-PD-1 monoclonal antibody tislelizumab in patients with unresectable HCC. Enrolled patients received tislelizumab (200 mg) intravenously every three weeks for a total of three or more 21-day treatment cycles, followed by long-term safety and survival assessments.

2.0.1 Patients

Patients were male and female adults (≥ 18 years of age), enrolled from international study sites, with histologically confirmed HCC that was not amenable to a curative treatment approach and who had received ≥ 1 line of systematic therapy for unresectable HCC. All patients were required to have an Eastern Cooperative Oncology Group (ECOG) performance status score of ≤ 1 [19].

2.0.2 Measures

HRQoL was assessed using three patient-reported outcome (PRO) instruments: the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30), the corresponding HCC-specific module (QLQ-HCC18), and the EQ-5D-5L. These PROs were collected at baseline and the first day of treatment cycle 2 (week 3), then every other treatment cycle up to cycle 12 (week 36). At each treatment cycle visit, the PRO administration occurred prior to any clinical activities or dosing. For purposes of this psychometric analysis, only QLQ-HCC18 and QLQ-C30 results are reported (the EQ-5D-5L was not employed in validation).

The EORTC QLQ-C30 [20] is a validated generic HRQoL instrument for cancer patients and comprises a global health status/QoL (GHS) scale (two items), five functional scales: physical functioning (five items), role functioning (two items), emotional functioning (four items), cognitive functioning (two items), social functioning (two items), as well as three symptom scales, and several single items: fatigue (three items), nausea and vomiting (two items), pain (two items), and dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial impact (one item each) [21]. The functional and symptom items are rated on a 4-point Likert scale (with 1 = ‘not at all’ to 4 = ‘very much’), while the GHS items are rated on a 7-point Likert scale (with 1 = ‘very poor’ to 7 = ‘excellent’). A high score on the GHS and functional scales indicates high HRQoL and a high level of functioning, whereas a high score on the symptom scales and items indicates a high level of symptom severity. The two individual GHS items were used as

concurrent validators. The GHS scale of the QLQ-C30 was used as the PRO anchor variable in test-retest reliability, ability to detect change, and meaningful within-patient change analyses.

The EORTC QLQ-HCC18 [22] measures HCC-specific symptoms and HRQoL. The instrument is an 18-item scale, consisting of six symptom scales and two single items: fatigue (three items), body image (two items), jaundice (two items), nutrition (five items), pain (two items), fever (two items), sexual interest (one item), and abdominal swelling (one item). Scores are based on a 4-point Likert scale (with 1 = 'not at all' to 4 = 'very much'); scaled scores for each domain ranged from 0-100 with a higher score indicating worse symptoms. In addition, an overall index score was calculated. Fatigue and index scores were prioritized in this validation exercise.

The ECOG performance status [19], a clinical measure of disease severity, was also used as a known-groups validator for this psychometric analysis. The ECOG criteria is used to assess how a patient's disease is progressing and the effect of the disease on a patient's activities of daily living and was assessed at the baseline visit.

In addition, demographic and medical history data, including age, gender, race, geographic region, line of therapy, and viral hepatitis infection status, were collected at the screening visit.

2.1 Statistical Analyses

In accordance with existing and emerging FDA guidance [17, 18], psychometric validation of the QLQ-HCC18 was conducted to measure the reliability (internal consistency and test-retest), construct validity (convergent validity and known-groups validity), ability to detect change, and MWPC. These analyses were conducted using the safety population, which included all patients receiving at least one dose of tislelizumab. Known-groups validity and MWPC analyses were stratified on several pre-defined subpopulations, including region (Asia [China/Taiwan] versus Europe), line of therapy (second-line versus third-line or greater), and viral hepatitis infection status (HBV/HCV positive versus hepatitis negative). Table 1 provides a summary of these analyses.

Table 1
Summary of psychometric analyses of QLQ-HCC18

Property	Analysis Period	Definition	Test	Success Criterion
Internal consistency	Baseline	Cronbach's α	No test, point estimate reported	$0.70 \leq \alpha$
Test-retest reliability	Baseline to week 3	ICC(2,1)	No test, point estimate reported	$0.70 \leq$ ICC(2,1)
Concurrent validity	Baseline	Spearman correlations	No test, point estimate reported	$\|0.40\| \leq r$
Known-groups validity	Baseline	Mean, mean difference, 95% CI, p-value, R^2 effect size	ANOVA	$p < 0.05$; effect size $\geq 5\%$
Ability to detect change	Baseline to week 9	Mean change from baseline in scores between anchor (QLQ-C30 GHS), 95% CI, p-value, and ω^2 semi-partial effect size	ANCOVA	$p < 0.05$; effect size $\geq 5\%$
Meaningful within-patient change	Baseline to week 9	Mean change from baseline in relation to change in anchor groups (QLQ-C30 GHS improvement, maintenance, deterioration) eCDFs and ePDFs plotted	No test, point estimates reported	No criterion, estimates reported

QLQ-HCC18: Quality of Life Questionnaire – Hepatocellular Carcinoma 18-question module; ICC: intraclass correlation coefficient; CI: confidence interval; ANOVA: analysis of variance; ANCOVA: analysis of covariance; QLQ-C30 GHS: Quality of Life Questionnaire – Core 30 global health status/QoL scale; eCDF: empirical cumulative distribution function; ePDF: empirical probability density function.

Descriptive statistics for continuous variables were reported as means, standard deviations (SDs), medians, and missing values. Descriptive statistics for categorical variables were reported as frequency counts and the percentage of patients in corresponding categories. Statistical significance was evaluated using a two-tailed $\alpha = 0.05$ level. Missing data for the QLQ-HCC18 and QLQ-C30 were handled according to the developer's manuals and no imputation was carried out [22, 23]. All analyses were performed using SAS (version 9.4) and R statistical software (version 3.6.1).

2.1.1 Reliability

Internal consistency evaluates score reliability by assessing the strength with which each item measures an assumed single domain. Internal consistency was assessed for each of the multi-item QLQ-HCC18 scales at baseline using Cronbach's alpha [24]. Internal consistency estimates of ≥ 0.70 were considered acceptable [23].

Test-retest reliability consists of measuring the degree to which an instrument is capable of reproducing scores across time in subjects whose condition has not changed [18]. Patients whose responses on the QLQ-C30 GHS scale anchor reflected no change in status between baseline and the first follow-up at week 3 were considered a stable subgroup and test-retest reliability was assessed for each of the QLQ-HCC18 scales and single items. In the

case of a continuous score, one appropriate measure of test-retest reliability is the two-way random intraclass correlation coefficient (ICC), employed in this analysis and denoted ICC(2,1) [25]. Test-retest reliability estimates of ≥ 0.70 indicate satisfactory reliability [26]. Both unconditional estimates and estimates conditioned on no change in GHS were estimated. Consistent with regulatory guidance, only estimates derived from the primary GHS anchor-based no-change definition (NC1, defined by GHS change score of 0 between baseline and week 3) are reported [17, 18, 27]. To limit the impact of possible treatment effects, three definitions of no change were examined in sensitivity analyses: unconditional, + 1 response category ('NC2'), or + 2 response categories ('NC3'). None of these definitions outperformed the pre-specified primary NC1 definition reported in this manuscript.

2.1.2 Construct Validity

Construct validity was assessed by tests of both concurrent validity and known-groups validity. Concurrent validity is a component of construct validity representing the extent to which two scales assessing similar constructs are related. This was estimated from Spearman correlations between the QLQ-HCC18 and QLQ-C30 scores at baseline. Larger positive correlations reflect convergent validity while small correlations or negative correlations reflect divergent or discriminant validity [28]. Spearman correlations of |0.40| or greater met the pre-specified criterion for acceptable concurrent validity [28].

Known-groups validity assesses whether PRO scores can be differentiated between clinically distinct groups. Known-groups validity was estimated for the QLQ-HCC18 scores at baseline. Known-groups validators included geographic region (Asia versus Europe), line of therapy (second-line versus third-line or greater), ECOG status (0 versus 1), and viral hepatitis infection status (HBV/HCV positive versus hepatitis negative). The difference in QLQ-HCC18 scores between each known-group was calculated and contrasted using analysis of variance (ANOVA), from which the mean difference between known-groups, corresponding 95% confidence interval (CI), p-value, and R-squared (R^2) effect size were estimated. Acceptable known-groups validity was achieved if a preponderance of the known-effect-groups had QLQ-HCC18 mean scores consistent with clinical expectations (i.e., more severe groups had worse symptoms or HRQoL compared to less severe groups). Such evidence was strengthened if and when the corresponding differences across known-groups were statistically significant and the corresponding R^2 was greater than 5%.

2.1.2 Ability to Detect Change

Ability to detect change is a facet of longitudinal validity that evaluates the relationship between changes in the PRO instrument of interest over time in the context of changes in external criteria (i.e., 'anchors') [29]. Ability to detect change was assessed by analyzing the extent to which QLQ-HCC18 change scores could be predicted by change in the QLQ-C30 GHS anchor variable. The QLQ-C30 GHS anchor groups were operationalized as follows: improvement was defined by > 0-point change from baseline to week 9; maintenance was defined as 0-point change from baseline to week 9; deterioration was defined as < 0-point change from baseline to week 9.

Analysis of covariance (ANCOVA) was used to estimate differences in QLQ-HCC18 change score marginal means across QLQ-C30 GHS anchor groups (improvement [effect] versus maintenance [reference]; deterioration [effect] versus maintenance [reference]), controlling for age, gender, region, and baseline QLQ-HCC18 mean. Effect size estimates were based on the Omega squared (ω^2) statistic [30]. Acceptable ability to detect change was pre-specified as estimates meeting the following criteria: significant differences ($p < 0.05$) in marginal means across anchor group contrasts and effect sizes exceeding 5%.

2.1.3 Meaningful Within-patient Change

Traditional estimation of meaningful change thresholds has relied on distribution and anchor-based methods. Increasingly, regulatory reviewers are emphasizing the latter, therefore anchor-based methods were the focus of the current analyses [27, 18, 17]. Furthermore, such estimates have emphasized between-group differences (e.g., minimally important differences or minimal clinically important differences). The FDA has justifiably taken the position that within-patient change is not acceptably approximated from between-group differences. Instead, regulatory guidance emphasizes MWPC for the derivation of clinical significance estimates [18].

Anchor-based methods aim to define the magnitude of MWPC on a PRO instrument of interest among patients classified as experiencing meaningful change (improvement/deterioration) on an 'anchor'. Anchor-based MWPC thresholds were obtained via calculation of mean change in QLQ-HCC18 scores from baseline to week 9 stratified on the QLQ-C30 GHS anchor groups described above. Known-groups validity was estimated for the QLQ-HCC18 scores at baseline. In addition to primary analyses based on the total sample, meaningful improvement estimates were stratified by geographic region (Asia versus Europe), line of therapy (second-line versus third-line or greater), and viral hepatitis infection status (HBV/HCV positive versus hepatitis negative). These estimates of mean change were then validated by visualizing differences in cumulative proportions achieving the point estimates stratified on anchor groups via empirical cumulative distribution functions (eCDFs) and empirical probability density functions (ePDFs).

3.0 Results

A total of 249 patients (138 second-line and 111 third-line or greater) were enrolled from 45 international centers in the BGB-A317-208 trial. The demographics and clinical characteristics of these patients are summarized in Table 2. This cohort had an average age of 60.3 years, was mostly male (87.1%), 50.6% Asian, balanced across ECOG status, and had an elapsed time from diagnosis to first dose of study drug of 38.7 months. Approximately a third of the patients (36.1%) were not HBV/HCV infected and approximately half were experiencing progressive disease (51.4%). The cohort had an average elapsed time from last systemic therapy dose to first study dose of 3.4 months. These patterns were similar across second-line and third-line or greater cohorts. A single patient who did not contribute QLQ-HCC18 data at baseline was excluded, leaving a final sample of 248 patients for the psychometric analyses.

Table 2
Patient demographics and clinical characteristics

Characteristic	Total Sample (N = 249) [1]	Line of Therapy	
		Second-line (n = 138)	Third-line or Greater (n = 111)
Age (years)			
Mean (SD)	60.3 (12.5)	60.2 (13.7)	60.4 (10.9)
Median	62.0	63.5	60.0
Min, Max	28, 90	28, 90	28, 82
Age group, n (%)			
<65 years	149 (59.8)	75 (54.3)	74 (66.7)
≥65 years	100 (40.2)	63 (45.7)	37 (33.3)
Gender, n (%)			
Male	217 (87.1)	121 (87.7)	96 (86.5)
Female	32 (12.9)	17 (12.3)	15 (13.5)
Race, n (%)			
Asian	126 (50.6)	74 (53.6)	52 (46.8)
Black or African American	4 (1.6)	2 (1.4)	2 (1.8)
White	96 (38.6)	43 (31.2)	53 (47.7)
Other	2 (0.8)	2 (1.4)	0 (0.0)
Not reported	21 (8.4)	17 (12.3)	4 (3.6)
ECOG performance status at baseline, n (%)			
0	129 (51.8)	70 (50.7)	59 (53.2)
1	120 (48.2)	68 (49.3)	52 (46.8)
Time from initial diagnosis to the first study dose (months)			
N	249	138	111
Mean (SD)	38.7 (39.6)	35.7 (37.8)	42.5 (41.6)
Median	24.9	21.4	28.1

SD: standard deviation; ECOG: Eastern Cooperative Oncology Group.

[1]A single patient who did not contribute QLQ-HCC18 data at baseline was excluded, leaving a final sample of 248 patients for the psychometric analyses.

Characteristic	Total Sample (N = 249) [1]	Line of Therapy	
		Second-line (n = 138)	Third-line or Greater (n = 111)
Min, Max	0.3, 269.6	2.3, 267.1	0.3, 269.6
Child-Pugh classification at baseline, n (%)			
A	248 (99.6)	138 (100.0)	110 (99.1)
B	1 (0.4)	0 (0.0)	1 (0.9)
Alpha-fetoprotein at baseline (ng/ml)			
>200 ng/mL	128 (51.4)	62 (44.9)	66 (59.5)
>400 ng/mL	112 (45.0)	53 (38.4)	59 (53.2)
Hepatitis virus infection, n (%)			
Uninfected	90 (36.1)	46 (33.3)	44 (39.6)
Hepatitis B only	123 (49.4)	71 (51.4)	52 (46.8)
Hepatitis C only	31 (12.4)	20 (14.5)	11 (9.9)
Coinfected	5 (2.0)	1 (0.7)	4 (3.6)
Number of lines of prior systemic therapy received, n (%)			
0	1 (0.4)	1 (0.7)	0 (0.0)
1	137 (55.0)	137 (99.3)	0 (0.0)
2	102 (41.0)	0 (0.0)	102 (91.9)
≥3	9 (3.6)	0 (0.0)	9 (8.1)
Best response to last systemic therapy, n (%)			
Complete response	0 (0.0)	0 (0.0)	0 (0.0)
Partial response	11 (4.4)	6 (4.3)	5 (4.5)
Stable disease	68 (27.3)	36 (26.1)	32 (28.8)
Progressive disease	128 (51.4)	71 (51.4)	57 (51.4)
Unknown/Not applicable/Missing	42 (16.9)	25 (18.1)	17 (15.3)
Time from the end of last systematic therapy to first dose (months)			
Mean (SD)	3.4 (6.8)	4.0 (8.6)	2.5 (3.0)

SD: standard deviation; ECOG: Eastern Cooperative Oncology Group.

[1]A single patient who did not contribute QLQ-HCC18 data at baseline was excluded, leaving a final sample of 248 patients for the psychometric analyses.

Characteristic	Total Sample (N = 249) [1]	Line of Therapy	
		Second-line (n = 138)	Third-line or Greater (n = 111)
Median	1.4	1.5	1.4
Min, Max	0.5, 79.0	0.5, 79.0	0.5, 17.7
SD: standard deviation; ECOG: Eastern Cooperative Oncology Group.			
[1]A single patient who did not contribute QLQ-HCC18 data at baseline was excluded, leaving a final sample of 248 patients for the psychometric analyses.			

3.1 Reliability

The Cronbach's alpha coefficients of three QLQ-HCC18 domains, namely fatigue, nutrition and index reflected acceptable internal consistency (0.71, 0.75, and 0.88, respectively). The remaining multi-item domains of body image, jaundice, pain, and fever did not display satisfactory internal consistency for this patient population (< 0.70).

Within the two assessments (baseline and 3-week follow-up) and across domains, 85–87 patients were included within the primary GHS-based no change (NC1) population upon which test-retest reliability was estimated. Test-retest reliability ICC(2,1) estimates indicated satisfactory reliability for six QLQ-HCC18 domains: fatigue, body image, nutrition, pain, sexual interest, and index (0.72, 0.70, 0.73, 0.75, 0.79, and 0.83 respectively). The remaining domains of jaundice, fever, and abdominal swelling did not display adequate test-retest reliability (< 0.70).

3.2 Construct Validity

Concurrent validity estimates are presented in Table 3. Correlations between QLQ-HCC18 scores and QLQ-C30 fatigue, nausea and vomiting, and pain domains met or exceeded the pre-specified criterion of $|0.40| \leq r$. The fatigue domain achieved this pre-specified criterion for 13 out of 16 (81.3%) correlations, whereas the index score achieved this pre-specified criterion for 15 out of 16 (93.8%) correlations. Conversely, there were weak correlations between non-relevant domains and items, suggesting acceptable discriminant validity. For example, the correlation between the QLQ-HCC18 fever domain and QLQ-C30 financial difficulties item was 0.21.

Table 3
Concurrent validity for the QLQ-HCC18 domains and the QLQ-C30 scores at baseline

QLQ-C30 Validators	QLQ-HCC18 Domains								
	Abdominal Swelling	Body Image	Fever	Fatigue	Jaundice	Nutrition	Pain	Sexual Interest	Index
Physical functioning	-0.46	-0.59	-0.36	-0.70	-0.27	-0.56	-0.56	-0.34	-0.71
Role functioning	-0.32	-0.55	-0.36	-0.62	-0.29	-0.51	-0.45	-0.35	-0.63
Emotional functioning	-0.40	-0.58	-0.41	-0.59	-0.33	-0.47	-0.53	-0.29	-0.64
Cognitive functioning	-0.22	-0.49	-0.43	-0.55	-0.30	-0.37	-0.44	-0.30	-0.56
Social functioning	-0.26	-0.44	-0.39	-0.55	-0.18	-0.51	-0.32	-0.46	-0.59
Fatigue	0.41	0.60	0.39	0.76	0.29	0.56	0.55	0.34	0.71
Nausea and vomiting	0.37	0.52	0.43	0.39	0.31	0.56	0.49	0.26	0.59
Pain	0.39	0.50	0.36	0.60	0.29	0.48	0.60	0.29	0.63
Dyspnoea	0.34	0.54	0.46	0.48	0.32	0.41	0.36	0.34	0.59
Insomnia	0.19	0.34	0.29	0.38	0.22	0.26	0.36	0.21	0.40
Appetite	0.29	0.50	0.35	0.49	0.31	0.68	0.44	0.20	0.57
Constipation	0.25	0.28	0.21	0.33	0.20	0.39	0.31	0.12	0.36
Diarrhoea	0.36	0.44	0.26	0.30	0.18	0.33	0.34	0.12	0.41
Financial difficulties	0.16	0.18	0.21	0.24	-0.02	0.23	0.15	0.38	0.32
GHS1	-0.33	-0.49	-0.34	-0.52	-0.22	-0.45	-0.44	-0.30	-0.56
GHS2	-0.34	-0.49	-0.31	-0.51	-0.19	-0.44	-0.41	-0.33	-0.56

QLQ-HCC18: Quality of Life Questionnaire – Hepatocellular Carcinoma 18-question module; QLQ-C30: Quality of Life Questionnaire – Core 30; GHS: global health status/QoL scale.

Known-groups validity of QLQ-HCC18 domains at baseline was defined upon geographic region, line of therapy, ECOG status, and viral hepatitis status. For the QLQ-HCC18 domains of fatigue, body image, jaundice and index, patients in Europe reported significantly higher mean scores compared with those patients in Asia. These mean differences were associated with effect sizes (R^2) indicating 2%, 5%, 4%, and 4% explained variance, respectively. For the QLQ-HCC18 body image domain, patients in the viral hepatitis negative group reported a significantly higher mean score (i.e., greater problems with body image) compared with those patients in the HBV/HCV positive group. For the QLQ-HCC18 jaundice domain, patients in the third-line or greater therapy group reported a significantly higher mean score compared with those patients in the second-line therapy group.

For all other domains and known-groups not mentioned above, the difference in known-group validators was not significant nor associated with meaningful explained variance at baseline. Domains for which no known-group demonstrated significance included the QLQ-HCC18 nutrition and fever domains, as well as the abdominal swelling and sexual interest items. Only 13.9% (5 out of 36) of known-groups analyses presented with mis-ordered means, wherein the known-group expected to have better symptoms and HRQoL demonstrated worse well-being.

3.2 Ability to Detect Change

Change scores were computed for the QLQ-HCC18 scores based on the QLQ-C30 GHS scale anchor groups of improvement, maintenance, and deterioration. The ability to detect change estimates are presented in Table 4. Clear differentiation of the QLQ-HCC18 change scores between improvement and maintenance groups were observed for body image, fatigue, pain, and index. Effect sizes were small (less than 0.10), most likely induced by the large variability in these data relative to the reasonable sample sizes, as indicated by the wide 95% CIs. No statistically significant changes were observed between improvement and maintenance groups for abdominal swelling, fever, jaundice, nutrition, and sexual interest. Clear differentiation of QLQ-HCC18 change scores between deterioration and maintenance groups were observed for fever and fatigue. No statistically significant differentiation was observed for the remaining QLQ-HCC18 symptom scores, including index.

Table 4
QLQ-HCC18 ability to detect change scores from baseline to week 9 by anchor group

QLQ-HCC18 Domain [1]	QLQ-C30 GHS Anchor [2]	Group Difference [3]	95% CI	P-value	Total Omega Effect Size	QLQ Omega Effect Size
Abdominal swelling	Improve (n = 50) vs. Maintenance (n = 63)	-5.98	-12.91, 0.95	0.090	0.036	0.017
	Deteriorate (n = 61) vs. Maintenance (n = 63)	2.39	-4.09, 8.87	0.466	-0.008	-0.004
Body image	Improve (n = 47) vs. Maintenance (n = 63)	-10.26	-16.55, -3.96	0.002	0.047	0.083
	Deteriorate (n = 63) vs. Maintenance (n = 63)	0.40	-4.03, 4.83	0.859	0.028	-0.008
Fever	Improve (n = 49) vs. Maintenance (n = 62)	-1.28	-4.75, 2.19	0.467	-0.006	-0.004
	Deteriorate (n = 63) vs. Maintenance (n = 62)	7.23	3.2, 11.25	0.001	0.002	0.094
Fatigue	Improve (n = 50) vs. Maintenance (n = 63)	-6.59	-12.65, -0.53	0.033	0.026	0.032
	Deteriorate (n = 61) vs. Maintenance (n = 63)	6.34	0.97, 11.72	0.021	-0.005	0.036
Jaundice	Improve (n = 50) vs. Maintenance (n = 63)	-2.90	-7.21, 1.4	0.184	0.015	0.007
	Deteriorate (n = 62) vs. Maintenance (n = 63)	-0.33	-4.48, 3.82	0.876	0.001	-0.008
Nutrition	Improve (n = 49) vs. Maintenance (n = 62)	-4.32	-9.07, 0.43	0.075	0.008	0.020
	Deteriorate (n = 61) vs. Maintenance (n = 62)	3.23	-0.36, 6.83	0.078	-0.008	0.018
Pain	Improve (n = 49) vs. Maintenance (n = 61)	-5.44	-10.73, -0.16	0.044	0.056	0.027
	Deteriorate (n = 63) vs. Maintenance (n = 61)	-0.14	-5.44, 5.16	0.958	-0.008	-0.008

QLQ-HCC18: Quality of Life Questionnaire – Hepatocellular Carcinoma 18-question module; QLQ-30 GHS: Quality of Life Questionnaire Cancer – Core 30 global health status/QoL scale; CI: confidence interval.

[1]QLQ-HCC18 domains are scored on a scale of 0-100 with higher scores indicate worse symptoms or more problems.

[2]QLQ-C30 GHS is scored on a scale of 1 to 7 with lower scores indicating reduced or low quality of life. Improve was defined as > 0-point change in QLQ-C30 GHS score; maintenance was defined as 0-point change; deterioration was defined as < 0-point change.

[3]Difference in marginal mean change score between anchors.

QLQ-HCC18 Domain [1]	QLQ-C30 GHS Anchor [2]	Group Difference [3]	95% CI	P-value	Total Omega Effect Size	QLQ Omega Effect Size
Sexual interest	Improve (n = 49) vs. Maintenance (n = 62)	-4.18	-13.89, 5.54	0.396	-0.007	-0.003
	Deteriorate (n = 60) vs. Maintenance (n = 62)	-1.14	-10.43, 8.16	0.809	-0.007	-0.008
Index	Improve (n = 50) vs. Maintenance (n = 64)	-5.31	-8.56, -2.05	0.002	0.071	0.078
	Deteriorate (n = 63) vs. Maintenance (n = 64)	2.23	-0.63, 5.09	0.125	-0.007	0.011
QLQ-HCC18: Quality of Life Questionnaire – Hepatocellular Carcinoma 18-question module; QLQ-30 GHS: Quality of Life Questionnaire Cancer – Core 30 global health status/QoL scale; CI: confidence interval.						
[1]QLQ-HCC18 domains are scored on a scale of 0-100 with higher scores indicate worse symptoms or more problems.						
[2]QLQ-C30 GHS is scored on a scale of 1 to 7 with lower scores indicating reduced or low quality of life. Improve was defined as > 0-point change in QLQ-C30 GHS score; maintenance was defined as 0-point change; deterioration was defined as < 0-point change.						
[3]Difference in marginal mean change score between anchors.						

3.3 Meaningful Within-patient Change

The point estimates for MWPC across anchor groups are presented for the total sample and stratified by region, line of therapy, and viral hepatitis infection status in Table 5. Within the primary (unstratified) analyses, point estimates for MWPC defining improvement were − 7.18 for QLQ-HCC18 fatigue and − 4.07 for QLQ-HCC18 index. Meaningful improvement estimates for the index scale stratified on either region or HBV/HCV infection were identical to the primary estimates. Region-stratified estimates of meaningful improvement for fatigue were within ± 1 point of the primary estimates. Line of therapy stratified estimates were within ± 2 of primary estimates for both fatigue and index. The viral hepatitis negative sample achieved greater fatigue improvement (-10) compared to the HBV/HCV infected sample (-5).

Table 5

QLQ-HCC18 meaningful within-patient change estimates from baseline to week 9 by anchor group

QLQ-HCC18 Domain [1]	QLQ-C30 GHS Anchor [2]	Mean Change						
		Total Sample	**Asia**	**Europe**	**Second-line Therapy**	**Third-line or Greater Therapy**	**Viral Hepatitis Negative**	**HBV/HCV Positive**
Abdominal swelling	Deteriorate	4.97 (n = 64)	6 (n = 32)	4 (n = 32)	3 (n = 30)	6 (n = 34)	8 (n = 24)	3 (n = 40)
	Improve	-2.66 (n = 50)	-5 (n = 19)	-1 (n = 31)	-3 (n = 30)	-2 (n = 20)	3 (n = 22)	-7 (n = 28)
	Maintenance	2.65 (n = 64)	-1 (n = 31)	6 (n = 33)	3 (n = 34)	2 (n = 30)	6 (n = 24)	1 (n = 40)
Body image	Deteriorate	2.92 (n = 64)	-2 (n = 32)	8 (n = 32)	-1 (n = 30)	6 (n = 34)	6 (n = 24)	1 (n = 40)
	Improve	-7.49 (n = 50)	-5 (n = 19)	-9 (n = 31)	-7 (n = 30)	-8 (n = 20)	-13 (n = 22)	-4 (n = 28)
	Maintenance	2.63 (n = 64)	0 (n = 31)	5 (n = 33)	3 (n = 34)	2 (n = 30)	6 (n = 24)	1 (n = 40)
Fever	Deteriorate	6.06 (n = 64)	4 (n = 32)	9 (n = 32)	4 (n = 30)	8 (n = 34)	9 (n = 24)	4 (n = 40)
	Improve	-2.39 (n = 50)	0 (n = 19)	-4 (n = 31)	-3 (n = 30)	-1 (n = 20)	-5 (n = 22)	-1 (n = 28)
	Maintenance	-1.13 (n = 64)	-3 (n = 31)	1 (n = 33)	-1 (n = 34)	-2 (n = 30)	4 (n = 24)	-4 (n = 40)
Fatigue	Deteriorate	5.34 (n = 64)	5 (n = 32)	6 (n = 32)	2 (n = 30)	9 (n = 34)	5 (n = 24)	5 (n = 40)
	Improve	-7.18 (n = 50)	-6 (n = 19)	-8 (n = 31)	-9 (n = 30)	-5 (n = 20)	-10 (n = 22)	-5 (n = 28)
	Maintenance	-0.87 (n = 64)	-4 (n = 31)	2 (n = 33)	0 (n = 34)	-2 (n = 30)	2 (n = 24)	-2 (n = 40)

QLQ-HCC18: Quality of Life Questionnaire – Hepatocellular Carcinoma 18-question module; QLQ-30 GHS: Quality of Life Questionnaire Cancer – Core 30 global health status/QoL scale; HBV/HCV: hepatitis B virus/hepatitis C virus.

[1]QLQ-HCC18 domains are scored on a scale of 0-100 with higher scores indicate worse symptoms or more problems.

[2]QLQ-C30 GHS is scored on a scale of 1 to 7 with lower scores indicating reduced or low quality of life. Improve was defined as > 0-point change in QLQ-C30 GHS score; maintenance was defined as 0-point change; deterioration was defined as < 0-point change.

eCDF: empirical cumulative distribution function; QLQ-HCC18: Quality of Life Questionnaire – Hepatocellular Carcinoma 18-question module.

Jaundice	Deteriorate	2.18 (n = 64)	2 (n = 32)	2 (n = 32)	2 (n = 30)	2 (n = 34)	2 (n = 24)	2 (n = 40)
	Improve	-0.06 (n = 50)	1 (n = 19)	-1 (n = 31)	-2 (n = 30)	2 (n = 20)	0 (n = 22)	0 (n = 28)
	Maintenance	2.57 (n = 64)	0 (n = 31)	5 (n = 33)	4 (n = 34)	0 (n = 30)	8 (n = 24)	-1 (n = 40)
Nutrition	Deteriorate	2.72 (n = 64)	1 (n = 32)	4 (n = 32)	2 (n = 30)	4 (n = 34)	3 (n = 24)	3 (n = 40)
	Improve	-4.67 (n = 50)	-3 (n = 19)	-6 (n = 31)	-7 (n = 30)	-1 (n = 20)	-6 (n = 22)	-3 (n = 28)
	Maintenance	-0.44 (n = 64)	-2 (n = 31)	1 (n = 33)	0 (n = 34)	-1 (n = 30)	2 (n = 24)	-2 (n = 40)
Pain	Deteriorate	2.3 (n = 64)	6 (n = 32)	-2 (n = 32)	2 (n = 30)	2 (n = 34)	-2 (n = 24)	5 (n = 40)
	Improve	-2.35 (n = 30)	-5 (n = 19)	-1 (n = 31)	-6 (n = 30)	3 (n = 20)	-1 (n = 22)	-4 (n = 28)
	Maintenance	2.44 (n = 64)	-1 (n = 31)	6 (n = 33)	5 (n = 34)	-1 (n = 30)	4 (n = 24)	1 (n = 40)
Sexual interest	Deteriorate	-1.73 (n = 64)	-3 (n = 32)	0 (n = 32)	5 (n = 30)	-7 (n = 34)	0 (n = 24)	-3 (n = 40)
	Improve	-4.78 (n = 50)	-5 (n = 19)	-4 (n = 31)	-3 (n = 30)	-7 (n = 20)	-5 (n = 22)	-5 (n = 28)
	Maintenance	-0.56 (n = 64)	0 (n = 31)	-1 (n = 33)	0 (n = 34)	-1 (n = 30)	-2 (n = 24)	0 (n = 40)
Index	Deteriorate	3.16 (n = 64)	2 (n = 32)	4 (n = 32)	2 (n = 30)	4 (n = 34)	4 (n = 24)	2 (n = 40)
	Improve	-4.07 (n = 50)	-4 (n = 19)	-4 (n = 31)	-5 (n = 30)	-2 (n = 20)	-5 (n = 22)	-4 (n = 28)
	Maintenance	1 (n = 64)	-1 (n = 31)	3 (n = 33)	2 (n = 34)	0 (n = 30)	4 (n = 24)	-1 (n = 40)

QLQ-HCC18: Quality of Life Questionnaire – Hepatocellular Carcinoma 18-question module; QLQ-30 GHS: Quality of Life Questionnaire Cancer – Core 30 global health status/QoL scale; HBV/HCV: hepatitis B virus/hepatitis C virus.

[1]QLQ-HCC18 domains are scored on a scale of 0-100 with higher scores indicate worse symptoms or more problems.

[2]QLQ-C30 GHS is scored on a scale of 1 to 7 with lower scores indicating reduced or low quality of life. Improve was defined as > 0-point change in QLQ-C30 GHS score; maintenance was defined as 0-point change; deterioration was defined as < 0-point change.

eCDF: empirical cumulative distribution function; QLQ-HCC18: Quality of Life Questionnaire – Hepatocellular Carcinoma 18-question module.

Within the primary (unstratified) analyses, point estimates for MWPC defining deterioration for QLQ-HCC18 fatigue and index were 5.34 and 3.16, respectively. In the case of the fatigue domain, estimates stratifying on either region

or HBV/HCV infection status were identical to the primary estimates (the one exception was Europe for which the estimate was 0.66 points higher). In the case of line of therapy, estimates were 2 and 9 respectively for second-line and third-line or greater, reflecting greater heterogeneity relative to the primary estimates. In the case of the index scale, all stratified estimates were within ± 1 of the primary estimates and therefore unaltered across population stratification.

The point estimates for MWPC for each anchor group definition were validated by eCDF and ePDF figures. In the case of meaningful improvement for fatigue domain scores, 60% of the improvement anchor group and 50% of the maintenance anchor group achieved the − 7.13 threshold, yielding a 10% improvement advantage. In the case of meaningful deterioration for fatigue scores, 38% of the deterioration anchor group and 18% of the maintenance anchor group achieved the 5.34 threshold, yielding a 20% advantage for maintenance. The eCDF for the QLQ-HCC18 fatigue score is presented in Fig. 1. The corresponding ePDF clarifies the overlap in fatigue domain change score distributions, but also demonstrates that the mass of distributions was offset as expected, with improvement skewed left, maintenance centered about a change score of zero, and deterioration skewed to the right.

4.0 Discussion

The present study examined the psychometric properties, namely reliability, construct validity, ability to detect change, and MWPC, of the EORTC QLQ-HCC18 instrument within the BGB-A317-208 trial population of patients with unresectable HCC. Within this population, evidence suggested that the QLQ-HCC18 demonstrates heterogenous psychometric properties. However, the QLQ-HCC18 fatigue and index domains were found to consistently demonstrate robust psychometrics.

With respect to reliability, this study found that only the QLQ-HCC18 fatigue, nutrition and index domains demonstrated acceptable internal consistency at baseline. This is not surprising given that previous validation studies found low alpha coefficients for the QLQ-HCC18 jaundice, pain, and fever domains, citing heterogeneity within the HCC patient population as the cause [7, 9, 12]. Specifically, these studies suggested heterogeneity of the items within the scales and within the patient population (e.g., region, viral hepatitis status) may be contributing factors. That may be the case, though a simpler explanation likely exists, and is reviewed within the limitations section. Acceptable test-retest reliability was found for fatigue, body image, nutrition, pain, sexual interest, and index. The observed low ICC estimates for the jaundice domain may have resulted from few patients presenting with jaundice upon admission to the trial.

Most convergent and discriminant validator correlations with the QLQ-HCC18 jaundice domain and sexual interest item failed to meet the pre-specified criterion defining acceptable concurrent validity. The fatigue domain achieved this pre-specified criterion for 13 of the 16 concurrent validity correlations, whereas the index domain achieved this pre-specified criterion for 15 of the 16 concurrent validity correlations. This was true for both convergent and discriminant validators.

Interpretable ability to detect change between patients improving versus maintaining according to the pre-specified QLQ-C30 GHS anchor thresholds was found for the fatigue, body image, pain, and index domain change scores. The same was found for ability to detect change between patients deteriorating versus maintaining for the fatigue domain. As expected, unbiased effect size estimates were low, indicating less than 10% explained variance across domains. This is often the case in oncology trials due to heterogeneity within the patient population, which increases dispersion, thereby attenuating effect-size magnitudes within the data. In this study, the estimated

anchor-based MWPC threshold defining clinical significance for the fatigue domain was found to be lower than previously reported within the literature [31, 32]. This may be due to the difference between the minimally important difference and MWPC frameworks. The revised MWPC deterioration estimates can be employed to define thresholds for progression endpoints, such as time to deterioration. The same is true for improvement endpoints, for which evidence was generated in this analysis indicating an ability of the QLQ-HCC18 fatigue domain to detect meaningful clinical improvement, which is a rare phenomenon in oncology PRO applications.

While the results of this study are important, they should be considered alongside some limitations. The most noteworthy limitation is that many of the QLQ-HCC18 domains did not consistently demonstrate optimal measurement properties in this HCC population. Specifically, body image, jaundice, pain, fever, and abdominal swelling did not display acceptable reliability. However, it is important to note that these domains consist of the fewest items within the QLQ-HCC18 instrument. Consistent with theory and previous evidence, the reliability of a score has been found to increase as the number of items contributing to the score increase [33, 34]. Additional limitations were related to validity and MWPC for domains other than fatigue and index. Jaundice and sexual interest failed to display acceptable validity. In addition, fever, nutrition, jaundice, abdominal swelling, and sexual interest did not show adequate ability to detect change.

Taken together, the validation evidence suggested that the QLQ-HCC18 fatigue and index domains consistently demonstrated robust psychometric properties. This appears to support the use of the fatigue and index domains as suitable patient-reported endpoints within an unresectable HCC population that had previously received one or more systemic therapies. Moreover, the ability to detect change and meaningful within-patient change analyses demonstrated that an uncommon degree of improvement was observed in this trial and the QLQ-HCC18 fatigue domain scores sensitively detected the effect of tislelizumab.

Declarations

Acknowledgements: The authors would like to thank Jason Allaire, PhD of Generativity Health Economics and Outcomes Research for assistance with the editing of this paper.

Funding: This study was funded by Beigene, Ltd.

Contributions: DS and LP were responsible for the analysis and interpretation of results. GB, JS and BT contribution to the interpretation of the study results. All authors participated in the writing and editing of the manuscript.

Conflict of Interest: GB, JS, and BT are employees of and own stock in BeiGene Ltd. DS and LP are employees of Pharmerit which received funding from BeiGene for this study,

References

1. **International Agency for Research on Cancer, World Health Organization. Cancer today (https://gco .iarc .fr/ today/ home).**
2. **El-Serag HB. Epidemiology of viral hepatitis and hepatocellular carcinoma. Gastroenterology. 2012;142(6):1264-73.e1. doi:10.1053/j.gastro.2011.12.061.**
3. **Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA: A Cancer Journal for Clinicians. 2015;65(2):87-108. doi:https://doi.org/10.3322/caac.21262.**

4. Ghouri Y, Mian I, Rowe J. Review of hepatocellular carcinoma: Epidemiology, etiology, and carcinogenesis. Journal of Carcinogenesis. 2017;16(1):1-. doi:10.4103/jcar.JCar_9_16.
5. Bosch FX, Ribes J, Cléries R, Díaz M. Epidemiology of hepatocellular carcinoma. Clinics in liver disease. 2005;9(2):191-211, v. doi:10.1016/j.cld.2004.12.009.
6. Jemal A, Ward EM, Johnson CJ, Cronin KA, Ma J, Ryerson B et al. Annual Report to the Nation on the Status of Cancer, 1975-2014, Featuring Survival. Journal of the National Cancer Institute. 2017;109(9). doi:10.1093/jnci/djx030.
7. Gandhi S, Khubchandani S, Iyer R. Quality of life and hepatocellular carcinoma. Journal of gastrointestinal oncology. 2014;5(4):296-317. doi:10.3978/j.issn.2078-6891.2014.046.
8. Bonnetain F, Paoletti X, Collette S, Doffoel M, Bouché O, Raoul JL et al. Quality of life as a prognostic factor of overall survival in patients with advanced hepatocellular carcinoma: results from two French clinical trials. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2008;17(6):831-43. doi:10.1007/s11136-008-9365-y.
9. Li L, Mo FK, Chan SL, Hui EP, Tang NS, Koh J et al. Prognostic values of EORTC QLQ-C30 and QLQ-HCC18 index-scores in patients with hepatocellular carcinoma - clinical application of health-related quality-of-life data. BMC cancer. 2017;17(1):8. doi:10.1186/s12885-016-2995-5.
10. Diouf M, Filleron T, Barbare JC, Fin L, Picard C, Bouché O et al. The added value of quality of life (QoL) for prognosis of overall survival in patients with palliative hepatocellular carcinoma. Journal of hepatology. 2013;58(3):509-21. doi:10.1016/j.jhep.2012.11.019.
11. Wible BC, Rilling WS, Drescher P, Hieb RA, Saeian K, Frangakis C et al. Longitudinal quality of life assessment of patients with hepatocellular carcinoma after primary transarterial chemoembolization. J Vasc Interv Radiol. 2010;21(7):1024-30. doi:10.1016/j.jvir.2010.03.005.
12. Chie WC, Blazeby JM, Hsiao CF, Chiu HC, Poon RT, Mikoshiba N et al. International cross-cultural field validation of an European Organization for Research and Treatment of Cancer questionnaire module for patients with primary liver cancer, the European Organization for Research and Treatment of Cancer quality-of-life questionnaire HCC18. Hepatology. 2012;55(4):1122-9. doi:10.1002/hep.24798.
13. Mikoshiba N, Tateishi R, Tanaka M, Sakai T, Blazeby JM, Kokudo N et al. Validation of the Japanese version of the EORTC hepatocellular carcinoma-specific quality of life questionnaire module (QLQ-HCC18). Health and quality of life outcomes. 2012;10:58. doi:10.1186/1477-7525-10-58.
14. Yang Z, Wan C, Li W, Cun Y, Meng Q, Ding Y et al. Development and Validation of the Simplified Chinese Version of EORTC QLQ-HCC18 for Patients with Hepatocellular Carcinoma. Cancer investigation. 2015;33(8):340-6. doi:10.3109/07357907.2015.1036280.
15. Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF et al. Safety, Activity, and Immune Correlates of Anti–PD-1 Antibody in Cancer. New England Journal of Medicine. 2012;366(26):2443-54. doi:10.1056/NEJMoa1200690.
16. Bersanelli M, Leonetti A, Buti S. The link between calcitriol and anticancer immunotherapy: vitamin D as the possible balance between inflammation and autoimmunity in the immune-checkpoint blockade. Immunotherapy. 2017;9(14):1127-31. doi:10.2217/imt-2017-0127.
17. Food and Drug Administration Guidance for Industry. Patient-reported outcome measures: use in medical product development to support labeling claims. Silver Spring; 2009. https://www.fda.gov/downloads/Drugs/.../Guidances/UCM193282.pdf.

18. Food and Drug Administration (FDA), 2018. Methods to Identify What is Important to Patients & Select, Develop or Modify Fit-for-Purpose Clinical Outcomes Assessments. https://www.fda.gov/media/116277/download.
19. Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. American journal of clinical oncology. 1982;5(6):649-55.
20. Aaronson NK, Ahmedzai S, Bergman 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. Journal of the National Cancer Institute. 1993;85(5):365-76. doi:10.1093/jnci/85.5.365.
21. Kemmler G, Holzner B, Kopp M, Dünser M, Margreiter R, Greil R et al. Comparison of two quality-of-life instruments for cancer patients: the functional assessment of cancer therapy-general and the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-C30. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 1999;17(9):2932-40. doi:10.1200/jco.1999.17.9.2932.
22. Blazeby JM, Currie E, Zee BC, Chie WC, Poon RT, Garden OJ. Development of a questionnaire module to supplement the EORTC QLQ-C30 to assess quality of life in patients with hepatocellular carcinoma, the EORTC QLQ-HCC18. European journal of cancer (Oxford, England : 1990). 2004;40(16):2439-44. doi:10.1016/j.ejca.2004.06.033.
23. Fayers PM, Machin D. Quality of Life: Assessment, Analysis and Interpretation. J Wiley & Sons Ltd, Chichester, 2000. ISBN: 0-471-96861-7. .
24. Cronbach L. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16(3):37.
25. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychological bulletin. 1979;86(2):420-8. doi:10.1037//0033-2909.86.2.420.
26. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J et al. Quality criteria were proposed for measurement properties of health status questionnaires. Journal of clinical epidemiology. 2007;60(1):34-42. doi:10.1016/j.jclinepi.2006.03.012.
27. Food and Drug Administration (FDA), 2014. Guidance for industry and FDA staff qualification process for drug development tools http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm230597.pdf.
28. Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale: Lawrence Erlbaum Associates, Publishers.
29. Liang MH. Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. Medical care. 2000;38(9 Suppl):li84-90.
30. Olejnik S, Algina J. Generalized eta and omega squared statistics: measures of effect size for some common research designs. Psychological methods. 2003;8(4):434-47. doi:10.1037/1082-989x.8.4.434.
31. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 1998;16(1):139-44. doi:10.1200/jco.1998.16.1.139.
32. King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 1996;5(6):555-67. doi:10.1007/bf00439229.
33. SPEARMAN C. CORRELATION CALCULATED FROM FAULTY DATA. British Journal of Psychology, 1904-1920. 1910;3(3):271-95. doi:https://doi.org/10.1111/j.2044-8295.1910.tb00206.x.

34. BROWN W. SOME EXPERIMENTAL RESULTS IN THE CORRELATION OF MENTAL ABILITIES1. British Journal of Psychology, 1904-1920. 1910;3(3):296-322. doi:https://doi.org/10.1111/j.2044-8295.1910.tb00207.x.

Figures

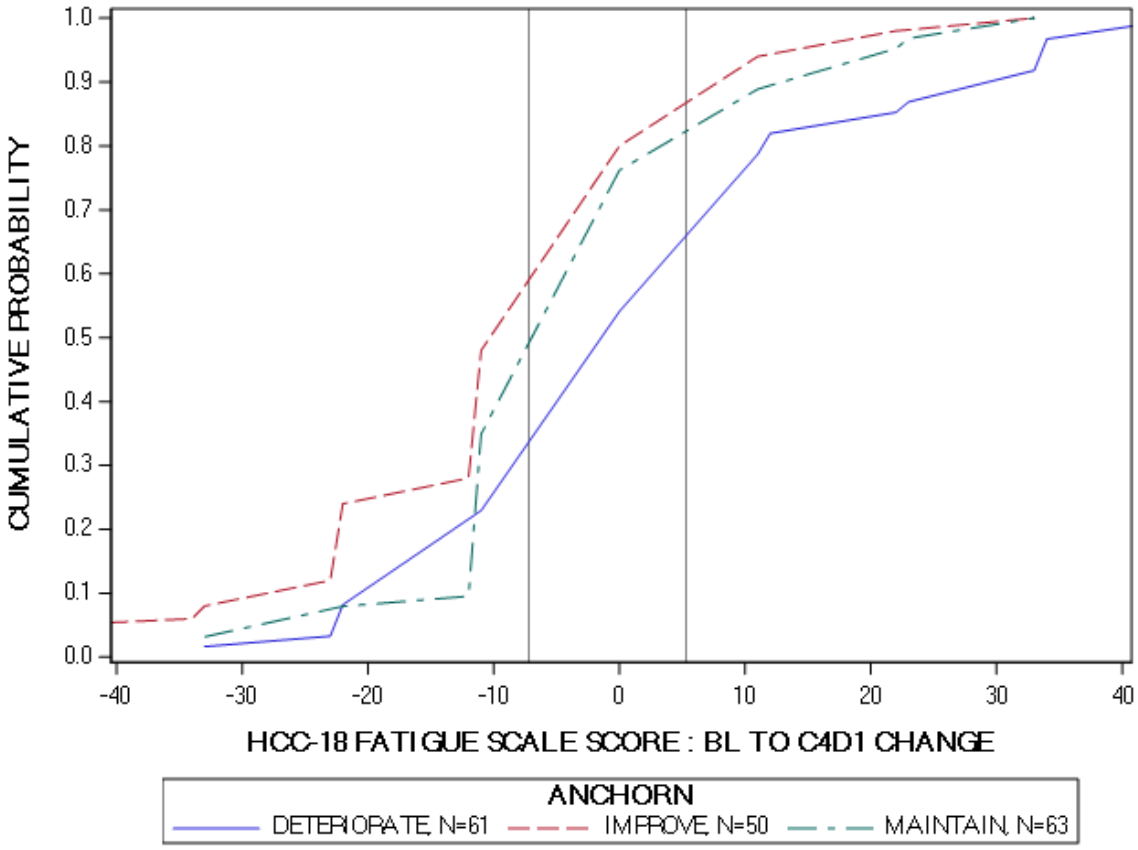

Figure 1

eCDF of QLQ-HCC18 fatigue domain change score from baseline to week 9 by anchor group. eCDF: empirical cumulative distribution function; QLQ-HCC18: Quality of Life Questionnaire – Hepatocellular Carcinoma 18-question module.