ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Patient-Reported Outcomes

Validation of the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 Summary Score in Patients With Hematologic Malignancies

Fabio Efficace, PhD,[1,*] Francesco Cottone, PhD,[1] Kathrin Sommer, PhD,[1] Jacobien Kieffer, PhD,[2] Neil Aaronson, PhD,[2] Peter Fayers, PhD,[3,4] Mogens Groenvold, MD, PhD,[5,6] Giovanni Caocci, MD,[7] Francesco Lo Coco, MD,[8] Gianluca Gaidano, MD,[9] Pasquale Niscola, MD, PhD,[10] Michele Baccarani, MD,[11] Gianantonio Rosti, MD,[11] Adriano Venditti, MD,[8] Emanuele Angelucci, MD,[12] Paola Fazi, MD,[1] Marco Vignetti, MD,[1] Johannes Giesinger, PhD[13]

[1]Italian Group for Adult Hematologic Diseases, Data Center and Health Outcomes Research Unit, Rome, Italy; [2]Department of Psychosocial Research, Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, The Netherlands; [3]Division of Applied Health Sciences, University of Aberdeen, Aberdeen, England, UK; [4]European Palliative Care Research Centre, Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway; [5]Department of Palliative Medicine, Bispebjerg Hospital, Copenhagen, Denmark; [6]Department of Public Health, University of Copenhagen, Copenhagen, Denmark; [7]Department of Medical Sciences, University of Cagliari, Cagliari, Italy; [8]Department of Biomedicine and Prevention, University of Rome Tor Vergata, Rome, Italy; [9]Division of Hematology, Department of Translational Medicine, University of Eastern Piedmont, Novara, Italy; [10]Hematology Unit, Sant'Eugenio Hospital, Rome, Italy; [11]Institute of Hematology "L. and A. Seràgnoli," Department of Experimental, Diagnostic and Specialty Medicine, "S. Orsola-Malpighi" University Hospital, University of Bologna, Italy; [12]IRCCS Ospedale Policlinico San Martino, Genova, Italy; [13]Psychiatry II, Medical University of Innsbruck, University Hospital Innsbruck, Innsbruck, Austria.

ABSTRACT

Objectives: We investigated the validity of the recently developed European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) summary score in patients with hematologic malignancies. Specifically, we evaluated the adequacy of a single-factor measurement model for the QLQ-C30, and its known-groups validity and responsiveness to change over time.

Methods: We used confirmatory factor analysis to test the single-factor model of the QLQ-C30, using baseline QLQ-C30 data (N = 2134). The QLQ-C30 summary score was compared to the original QLQ-C30 scales using general (age, sex, Eastern Cooperative Oncology Group performance status, comorbidity) and disease-specific (red blood cell transfusion dependency) groups. Repeated measurements allowed us to investigate responsiveness to change in a subgroup of patients with acute myeloid leukemia.

Results: The single-factor model of the QLQ-C30 exhibited adequate fit in patients with hematologic malignancies. Known-group comparisons generally supported the construct validity of the summary score when using more general grouping variables (sociodemographics, broad clinical parameters). Nevertheless, when groups were formed on the basis of disease-specific variables (eg, transfusion dependency), the summary score performed less well the some of the original, separate scales of the QLQ-C30.

Conclusion: Our findings provide support for the validity of the single-factor model of the EORTC QLQ-C30 in patients with hematologic malignancies. Specifically, the results suggest that the summary score can be used as an endpoint in this population when symptom- or other health domain-specific hypotheses are not available.

Keywords: confirmatory factor analysis, EORTC QLQ-C30, hematologic malignancies, leukemia, summary score, validation.

VALUE HEALTH. 2019; 22(11):1303–1310

Conflicts of interest: Emanuele Angelucci: *Honoraria* from Novartis and Celgene; *involvement in local advisory board* for Jazz Pharma and Roche; *participation in DMC* for Celgene, Vertex Pharma Inc and Crispr Therapeutics. Michele Baccarani: *Honoraria* from Incyte, Pfizer e Novartis. Fabio Efficace: *Consultancy* for Bristol-Myers Squibb, Amgen, Orsernix, Incyte and Takeda. Gianluca Gaidano: *advisory board* for Janssen, Astra-Zeneca, Abbvie and Sunesis; *speaker bureau* Janssen and Abbvie. Gianantonio Rosti: *speaker bureau* for Novartis, BMS, Pfizer, Incyte; *research support* for Novartis, Pfizer, Incyte.

* Address correspondence to: Fabio Efficace, PhD, Health Outcomes Research Unit, Italian Group for Adult Hematologic Diseases, GIMEMA Data Center, Via Benevento, 6, Rome 00161, Italy. Email: f.efficace@gimema.it

https://doi.org/10.1016/j.jval.2019.06.004

Introduction

The European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30)[1] is one of the most widely used patient-reported outcome (PRO) measures in assessing health-related quality of life (HRQOL), functional health, and symptoms in cancer clinical trials.[2] The QLQ-C30 assesses 15 health domains, including physical and emotional functioning and key cancer symptoms such as fatigue, pain, dyspnea, and nausea. Although the wide range of scales included in the QLQ-C30 provides a comprehensive overview of a patient's HRQOL, it also presents statistical challenges in clinical trials due to multiple testing and the resulting risk of type I errors.

Although several approaches are available for correcting or compensating for multiple testing, often all QLQ-C30 scales are analyzed without such corrections.[3] Sometimes, to avoid the problem, the 2-item global health status/quality of life (QOL) scale is used as a primary, single endpoint. Nevertheless, the global health status/QOL scale of the QLQ-C30 may not be sensitive enough to capture the relevant health-specific aspects for a given cancer population under study. For example, in phase III randomized controlled trials, the selection of the primary HRQOL endpoint should ideally be based on previous empirical data of specific health domains most likely to be affected by the treatment being tested.[4-6] Nevertheless, this information might not be available from previous studies, or the selection of specific health domains as primary outcomes may be inappropriate, such as in research settings using large heterogeneous cancer population-based registries.

To overcome these challenges, the QLQ-C30 Summary Score has been introduced as an additional scoring algorithm.[7] This score is based on the statistical evaluation of 7 higher-order models for the 15 health domains of the QLQ-C30 that were defined on the basis of relevant literature and expert opinion.[8] The models comprised either a single higher-order HRQOL factor or different combinations of 2 higher-order factors (eg, physical and mental health, physical burden and mental function, symptom burden and function). In a replication study,[7] all models were re-evaluated using data from 3282 patients with mixed cancer diagnoses. Based on model fit and parsimony, the single-factor model was selected to serve as the basis for the QLQ-C30 Summary Score. The QLQ-C30 Summary Score was also shown to be as discriminative as the best-performing single scales of the QLQ-C30 regarding tumor stage, performance status, and change over time. In line with these results, a recent study comparing a heterogeneous sample of 4020 cancer patients with a general population sample also demonstrated that the summary score was more discriminative than any single scale of the QLQ-C30.[9] Although these studies provide support for the use of the QLQ-C30 Summary Score, the data sets used in these studies comprised heterogeneous samples of cancer patients with mainly solid tumors.

Given the outstanding progress made in clinical research, hematologic cancers are now being treated with newly adopted disease management strategies, such as molecularly targeted therapies and monoclonal antibodies. This has resulted in substantially improved survival rates. Life expectancy of patients with some types of leukemia now approaches that of their peers in the general population,[10] and targeted therapies in some types of acute leukemia have contributed to remarkable survival improvements.[11] This has led to a paradigm shift within the hematologic research field towards assuming a more chronic therapy model for disease, making the argument for implementation of PRO measures more compelling.[12-15] In this context, it is relevant to investigate the psychometric performance of the QLQ-C30 Summary Score in patients with hematologic malignancies

Our study objectives were to investigate (1) whether the previously established single-factor model underlying the QLQ-C30 Summary Score could be replicated in patients with hematologic malignancies, and (2) the validity of this summary score using known-group comparisons and analysis of responsiveness to change over time using both general and disease-specific anchors.

Patients and Methods

Data Source

For the purpose of the current study, we formed a pooled database of 6 individual datasets. These datasets comprise cross-sectional and longitudinal studies conducted by Gruppo Italiano Malattie Ematologiche dell'Adulto (GIMEMA) in patients diagnosed with acute myeloid leukemia (AML) (n = 1),[16] chronic myeloid leukemia (n = 2),[17-19] acute promyelocytic leukemia (n = 1),[4,11] and myelodysplastic syndromes (MDS) (n = 2).[20-23] The pooled database for the purpose of this analysis consisted of 2134 patients who completed the QLQ-C30 at baseline and, only for patients with AML, after the induction treatment phase (n = 255). Most patients were from Italy (90.2%). Sociodemographic and clinical variables, including age, sex, Eastern Cooperative Oncology Group (ECOG) performance status, comorbidity, and transfusion dependency, were available for nearly all datasets.

EORTC QLQ-C30

The EORTC QLQ-C30[1] is a cancer-specific questionnaire with 30 items organized into 15 scales. There are 5 multi-item functional scales (physical [PF], role [RF], cognitive, emotional [EF], and social [SF]); 3 multi-item symptom scales (fatigue [FA], nausea and vomiting, and pain); 6 single-item symptom scales (dyspnea [DY], insomnia, appetite loss, constipation [CO], diarrhea [DI], and financial difficulties [FI]); and a 2-item health status/global QOL scale. All questions are answered on a 4-point Likert-type scale, with the exception of the 2 items of the QOL scale that use a 7-point scale. For the functioning and the QOL scales, a higher score indicates better functioning. For the symptoms scales, a higher score indicates a higher level of symptom burden. The QLQ-C30 summary score was calculated as the mean of the combined 13 QLQ-C30 scale scores (excluding FI and QOL). Prior to calculating the mean, symptom scale scores are reversed so that higher scores indicate lower symptom burden.[7]

Statistical Analysis

To replicate the previously established single-factor measurement model underlying the QLQ-C30 Summary Score, we used the same modeling procedure as in Gundy et al[8] and Giesinger et al[7]. In brief, the single-factor model was estimated by using confirmatory factor analysis (CFA). The QLQ-C30 scales were modeled as first-order factors (except for the FI scale, which was excluded from analysis). Then all first-order factors apart from QOL were loaded on the second-order HRQOL factor (the summary score). The QOL factor was left free to correlate with the higher-order HRQOL factor. To ensure model identification, one of the item-loadings for each factor and the variance of the second-order HRQOL factor were set to 1. The residual variance of the single-item factors was fixed at 20% of the total variance. We used the

weighted least-squares estimator with adjustment for means and variances, and missing data were handled using pairwise deletion. In accordance with previous assumptions,[7] standardized factor loadings were expected to be greater than 0.4. The adequacy of the CFA models was evaluated using the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), and the root mean square error of approximation (RMSEA) to assess goodness of fit of the single-factor model.[24] Values of CFI and TLI above 0.95 or 0.90 indicate good or acceptable fit, respectively. For RMSEA, values below 0.05 indicate good fit and below 0.08 indicate acceptable fit.[25]

After confirming the single-factor HRQOL measurement model in this sample, we employed univariate linear regression analysis to investigate the known-groups validity (ie, the ability to distinguish between groups known to be different on certain background characteristics) of the QLQ-C30 Summary Score and the original, separate scales of the QLQ-C30. General linear mixed-model analysis was used to compare longitudinal changes in HRQOL. General groups were formed according to ECOG performance status (0-1 vs 2-3), comorbidity (yes vs no), sex (male vs female), and age ($\leq$ 60 vs $>$ 60 years). Disease-specific groups were formed on the basis of red blood cell transfusion dependency groups (yes vs no) for MDS patients, and baseline versus post-induction scores for AML patients only. This latter comparison was used to evaluate the responsiveness to change.

Mean difference/change in the QLQ-C30 Summary Score and the original QLQ-C30 scale scores were accompanied by Cohen's d effect sizes (ESs), with an ES of 0.2 considered small, 0.5 moderate, and 0.8 large.[26] To compare the performance of each QLQ-C30 scale to that of the QLQ-C30 Summary Score, we also computed relative validity (RV) estimates from the known-groups regression analyses. The RV is the ratio of the F statistics from the comparator and the reference scale (F_{comp}: F_{ref}).[26] In this study, the QLQ-C30 Summary Score was used as the reference scale. RV value lower than 1 indicates a better performance of the QLQ-C30 Summary Score, while a RV value greater than 1 indicates a better performance of the original QLQ-C30 scale scores. We used the lavaan package from R software, version 3.5.1,[27] to perform CFA. All other analyses were performed using SAS software, version 9.4 (SAS Institute Inc).

Results

Patient Characteristics

Analyses were based on data from 2134 patients diagnosed with AML (n = 477), acute promyelocytic leukemia (n = 244), chronic myeloid leukemia (n = 382), and MDS (n = 1031). Overall, the mean age of patients was 51.5 years (interquartile range: 41-60 years), and 56.4% and 43.6% were male and female patients, respectively. At least 1 comorbid condition was reported by 59.9% of the patients, and most patients had an ECOG status 0 (47.4%) or 1 (39.3%). In the MDS group, 29% of the patients were dependent on red blood cell transfusions. Further details are reported in Table 1.

Replication of the Higher-Order Single-Factor Model

The single-factor model including all QLQ-C30 scales (except FI) showed an acceptable to good model-data-fit with an RMSEA of 0.064 (90% confidence interval [CI] 0.063-0.066), a CFI of 0.963, and a TLI of 0.959. All standardized factor loadings exceeded the predefined threshold of 0.40, with the exception of CO (0.337) and DI (0.357). The bivariate correlation between the QLQ-C30 Summary Score and the QOL scale was r = 0.72. Further details are given in Figure 1.

Table 1. Patient characteristics (N = 2134).

Variables	Value
Age (years)	
Mean (SD)	51.5 (14.5)
Median (IQR)	51.0 (41.0-60.0)
Sex, n (%)	
Male	1196 (56.4)
Female	925 (43.6)
Missing	13
ECOG, n (%)*	
0	885 (47.4)
1	732 (39.3)
2	215 (11.5)
3	33 (1.8)
Missing	25
Comorbidity, n (%)†	
No	664 (40.1)
Yes	991 (59.9)
Missing	2
Diagnosis, n (%)	
AML	477 (22.4)
APL	244 (11.4)
CML	382 (17.9)
MDS	1031 (48.3)
Transfusion dependency, n (%)‡	
No	721 (71.0)
Yes	295 (29.0)
Missing	15
Country, n (%)	
Italy	1926 (90.2)
Germany	50 (2.4)
Austria	39 (1.8)
China	27 (1.3)
Czech Republic	25 (1.2)
United Kingdom	20 (0.9)
Croatia	13 (0.6)
Belgium	10 (0.5)
France	10 (0.5)
United States	8 (0.4)
Brazil	6 (0.3)

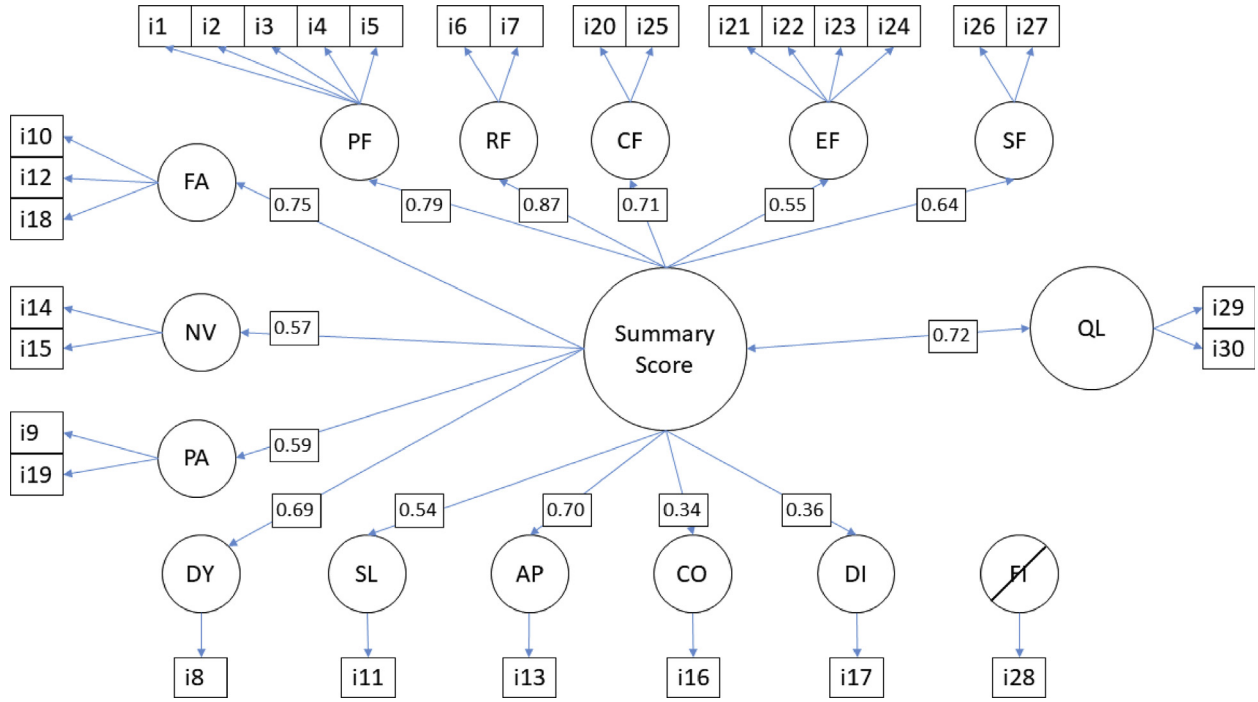
AML indicates acute myeloid leukemia; APL, acute promyelocytic leukemia; CML, chronic myeloid leukemia; ECOG, Eastern Cooperative Oncology Group; IQR, interquartile range; MDS, myelodysplastic syndromes.
*Available in 5 of 6 datasets (N = 1890).
†Available in 5 of 6 datasets (N = 1657).
‡Considered only for MDS patients (N = 1031).

General Known-Group Comparisons

When comparing patients with high and low ECOG status (0-1 vs 2-3), the QLQ-C30 Summary Score outperformed 12 of 14 QLQ-C30 scales, with a larger ES (0.66) and better performance in relative validity (see Table 2). With the QLQ-C30 Summary Score as reference scale, 8 of the QLQ-C30 scales even showed an RV lower than 0.50, with corresponding ES ranging from 0.12 to 0.41. Only the PF scale (ES = 0.80, RV = 1.47) and the RF scale (ES = 0.69, RV = 1.08) outperformed the QLQ-C30 Summary Score. When comparing patients with and without comorbidities, the QLQ-C30 Summary Score performed better than 11 of the single QLQ-C30 scales, with ES = 0.17 (Table 3). Three scales outperformed the QLQ-C30 Summary Score; PF (ES = 0.19, RV = 1.27), pain (ES = −0.19, RV = 1.27), and FA (ES = −0.18, RV = 1.04).

The comparisons of men and women yielded a difference in ES of 0.25 for the QLQ-C30 Summary Score, with PF (ES = 0.26,

Figure 1. Single-factor model for the EORTC QLQ-C30 with factor loadings.



AP indicates appetite loss; CF, cognitive function; CO, constipation; DI, diarrhea; DY, dyspnea; EF, emotional function; EORTC QLQ-C30, European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30; FA, fatigue; FI, financial impact; NV, nausea and vomiting; PA, pain; PF, physical function; QL, global quality of life; RF, role function.

RV = 1.05) and EF (ES = 0.25, RV = 1.01) being the only QLQ-C30 scales outperforming the QLQ-C30 Summary Score. For all other scales, effect sizes were below ES = 0.22. When using age as the grouping variable, the difference between patients above versus below 60 years was largest for PF (ES = 0.31, RV = 1.63) followed by the QLQ-C30 Summary Score (ES = 0.17). For all other scales, the ES was 0.14 or lower. The QLQ-C30 Summary Score outperformed the remaining 13 QLQ-C30 scales.

Table 2. Effect sizes and relative validities for the EORTC QLQ-C30 Summary Score and the individual scales using ECOG performance status as grouping variable.

QLQ-C30 Scale	ECOG 0 and 1, N = 1617		ECOG 2 and 3, N = 248		RV	ES
	Mean	SD	Mean	SD		
Summary score	80.3	14.6	70.3	18.3	1.00	0.66
Physical functioning	77.9	20.7	60.6	26.7	1.47	0.80
Role functioning	74.7	28.6	54.7	31.8	1.08	0.69
Cognitive functioning	84.3	19.4	76.0	26.0	0.37	0.40
Emotional functioning	74.3	21.8	70.7	24.9	0.06	0.16
Social functioning	81.3	25.2	67.8	30.7	0.61	0.52
Global QOL	59.6	23.6	46.1	23.3	0.75	0.57
Fatigue	36.3	25.8	52.6	27.2	0.90	−0.63
Nausea/vomiting	7.2	14.9	10.6	18.1	0.11	−0.22
Pain	18.1	24.1	27.2	27.6	0.32	−0.37
Dyspnea	24.9	27.4	38.6	32.6	0.54	−0.49
Insomnia	23.7	28.3	30.5	32.5	0.13	−0.24
Appetite loss	16.1	26.3	27.3	32.3	0.39	−0.41
Constipation	15.5	24.8	20.5	29.6	0.09	−0.20
Diarrhea	8.2	18.3	10.5	22.0	0.04	−0.12

ECOG indicates Eastern Cooperative Oncology Group; ES, effect size; QLQ-C30, Quality of Life Questionnaire Core 30; QOL, quality of life; RV, relative validity.

Table 3. Effect sizes and relative validities for the EORTC QLQ-C30 Summary Score and the individual scales using comorbidity as grouping variable.

QLQ-C30 Scale	No comorbidity, N = 664		At least 1 comorbidity, N = 991		RV	ES
	Mean	SD	Mean	SD		
Summary score	82.6	14.3	80.0	16.0	1.00	0.17
Physical functioning	79.0	21.0	74.7	22.7	1.27	0.19
Role functioning	79.5	26.2	75.7	29.2	0.63	0.14
Cognitive functioning	84.8	19.8	82.2	20.7	0.55	0.13
Emotional functioning	76.6	21.8	74.8	21.4	0.23	0.08
Social functioning	83.8	24.9	82.2	24.8	0.13	0.06
Global QOL	64.1	23.6	62.2	22.4	0.25	0.09
Fatigue	31.5	25.6	36.0	26.2	1.04	−0.18
Nausea/vomiting	5.6	12.9	6.4	14.2	0.10	−0.06
Pain	14.9	22.4	19.5	24.5	1.27	−0.19
Dyspnea	23.1	26.4	25.0	29.2	0.16	−0.07
Insomnia	22.0	27.2	24.8	29.2	0.31	−0.10
Appetite loss	11.5	22.1	14.7	26.2	0.58	−0.13
Constipation	14.8	24.0	15.9	25.9	0.06	−0.04
Diarrhea	6.5	15.6	9.6	20.4	0.92	−0.17

EORTC QLQ-C30 indicates European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30; ES, effect size; QOL, quality of life; RV, relative validity.

Disease-Specific Known-Group Comparisons

Data on transfusion dependency were available for the 2 datasets with MDS patients, as this is a critical issue for the clinical management of these patients. For all patients, transfusion dependency was defined prior to baseline HRQOL assessment, based on previously established criteria for this cancer population.[28] The largest differences between those dependent on transfusion versus those who were not were observed for QOL (ES = 0.33), DY (ES = −0.29), and RF (ES = 0.27), and the smallest differences for CO (ES = 0.02) and DI (ES = −0.02) (Table 4). The effect size for the QLQ-C30 Summary Score was ES = 0.21. Six scales had an RV

Table 4. Effect sizes and relative validities for the EORTC QLQ-C30 Summary Score and the individual scales using dependency on RBC transfusion as grouping variable in MDS patients.

QLQ-C30 Scale	No RBC transfusion dependency, N = 721		RBC transfusion dependency, N = 295		RV	ES
	Mean	SD	Mean	SD		
Summary score	80.2	15.1	77.1	15.3	1.00	0.21
Physical functioning	74.4	22.2	68.6	22.0	1.56	0.26
Role functioning	75.0	28.9	67.2	30.0	1.67	0.27
Cognitive functioning	83.1	20.6	81.7	20.7	0.11	0.07
Emotional functioning	74.2	21.9	75.4	22.1	0.07	−0.05
Social functioning	82.4	25.3	75.1	28.3	1.79	0.28
Global QOL	60.3	22.6	52.8	22.9	2.54	0.33
Fatigue	35.9	26.0	42.1	25.8	1.32	−0.24
Nausea/vomiting	5.2	12.6	6.2	13.1	0.15	−0.08
Pain	17.2	23.5	20.6	25.9	0.46	−0.14
Dyspnea	25.4	27.8	33.7	31.5	1.90	−0.29
Insomnia	24.8	28.7	22.2	27.9	0.18	0.09
Appetite loss	14.4	25.4	19.0	27.3	0.75	−0.18
Constipation	16.5	26.4	16.1	26.2	0.01	0.02
Diarrhea	6.9	17.2	7.3	17.4	0.01	−0.02

EORTC QLQ-C30 indicates European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30; ES, effect size; MDS, myelodysplastic syndromes; QOL, quality of life; RBC, red blood cell; RV, relative validity.

Table 5. Effect sizes and relative validities for the EORTC QLQ-C30 Summary Score and the individual scales in a comparison of patients with AML at diagnosis and post-induction chemotherapy (treatment responders only).

QLQ-C30 Scale	At diagnosis, N = 255		Post-induction chemotherapy, N = 255		RV	ES
	Mean	SD	Mean	SD		
Summary score	76.7	16.0	83.3	12.4	1.00	0.41
Physical functioning	79.7	21.9	80.3	17.1	0.01	0.03
Role functioning	64.3	33.1	68.0	30.1	0.07	0.11
Cognitive functioning	84.8	19.7	89.8	15.1	0.43	0.26
Emotional functioning	72.0	22.6	82.2	17.2	1.54	0.45
Social functioning	73.3	27.7	73.9	26.2	<0.01	0.02
Global QOL	49.3	26.3	65.5	21.0	2.23	0.61
Fatigue	44.2	28.6	32.1	21.2	1.10	−0.42
Nausea/vomiting	9.9	17.0	8.0	16.0	0.05	−0.12
Pain	20.1	26.0	11.6	19.2	0.60	−0.33
Dyspnea	29.0	29.2	10.8	19.2	2.32	−0.62
Insomnia	27.2	30.9	16.7	22.9	0.63	−0.34
Appetite loss	25.4	29.9	14.6	24.3	0.56	−0.36
Constipation	15.3	25.4	10.3	20.5	0.19	−0.20
Diarrhea	5.6	14.8	7.2	16.6	0.04	0.11

AML indicates acute myeloid leukemia; EORTC QLQ-C30, European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30; ES, effect size; QOL, quality of life; RV, relative validity.

indicative of better performance than the QLQ-C30 Summary Score: QOL (RV = 2.54), DY (RV = 1.90), SF (RV = 1.79), RF (RV = 1.67), PF (RV = 1.56), and FA (RV = 1.32). For the remaining 8 scales, the RV was lower than 1, indicating poorer performance than the QLQ-C30 Summary Score.

When comparing the 255 AML patients before and after induction chemotherapy, we found the largest mean change for DY (ES = −0.62), QOL (ES = 0.61), and EF (ES = 0.45), and the smallest change for SF (ES = 0.02) and PF (ES = 0.03). For the QLQ-C30 Summary Score the mean change was ES = 0.41. Based on the RVs, the scales DY (2.32), QOL (2.23), EF (1.54), and FA (1.10) outperformed the QLQ-C30 Summary Score. Details are provided in Tables 4 and 5.

Discussion

Our results confirm the adequacy of a single-factor measurement model of the QLQ-C30 in patients with hematologic malignancies. Known-group comparisons indicated that the QLQ-C30 Summary Score derived from this model is consistent among the most discriminatory scales for generic patient characteristics such as ECOG performance status, comorbidity, age, and sex. A few individual scales (PF in particular) outperformed the QLQ-C30 Summary Score in specific comparisons, but no single scale outperformed the summary score across all of the generic comparisons.

In the 2 comparisons specific to hematologic patients, the QLQ-C30 Summary Score was less discriminatory. It was only the seventh most discriminatory scale in the comparison of patients with or without dependency on blood transfusions, and the fifth most discriminatory scale in responsiveness to change from pre- to post-induction chemotherapy in AML patients. In these 2 comparisons, 2 specific symptom scales, DY and FA, both key symptoms in AML patients that have been shown to be sensitive to treatment changes,[29] outperformed the QLQ-C30 Summary Score. This was also true for the global QOL scale and EF, which have been shown to be sensitive to changes over time in previous AML longitudinal studies.[30]

Our results are largely in line with the previous study by Giesinger et al[7] that relied primarily on data from patients with solid tumors. Model-data fit statistics were very similar.[7] Only the standardized factor loadings for constipation and diarrhea in the current study were somewhat lower than in the previous study.[7] This may reflect the lesser importance of these symptoms in patients with hematologic malignancies, as was also evidenced by the results of the known-group comparisons. Our results for the more generic group comparisons were also largely similar to those of Giesinger et al,[7] where the QLQ-C30 Summary Score generally outperformed the individual scales. Disease-specific variables were not included in the Giesinger et al study, due to the heterogeneity of their cancer sample, and thus comparisons with the current results are not possible. Our finding that the presence of comorbid conditions is most strongly associated with PF is in line with previous studies. Mols et al[31] used reference data from the Dutch general population to compare participants with and without comorbid conditions. In that study, the QLQ-C30 Summary Score showed the second largest group difference, with only the PF scale being more discriminative.

The good performance of the global QOL scale observed for our disease-specific comparisons in AML and MDS patients differs from that seen in previous studies in patients with solid tumors,[7,32,33] which questioned the sensitivity of the global QOL scale. Although this finding warrants further research, we believe that, overall, our results support the strategy of employing the QLQ-C30 Summary Score as a study endpoint in cases where previous knowledge does not allow for domain-specific

hypotheses or when the sample being studied includes mixed hematologic populations. Where domain-specific hypotheses can be posed, specific QLQ-C30 scale scores may be more appropriate as primary HRQOL endpoints.

Our study has several limitations that should be noted. Our datasets included mostly patients from centers in Italy and did not provide the variety of countries in the original validation study of the QLQ-C30 Summary Score.[7] Also, data for comorbidity and ECOG performance status were not available for all datasets.

A strength of our study is the large dataset including patients from various multicenter trials and observational studies. Also, our study represents a truly independent replication of the single-factor model, whereas the previous evaluations of model-data fit[7,8] relied on 2 random halves of the same datasets. Finally, although not covering all hematologic malignancies, our data included a wide spectrum of leukemia subtypes, including acute and chronic leukemia conditions as well as patients with myelodysplastic syndromes.

As discussed above, the QLQ-C30 Summary Score has been introduced to overcome problems with multiplicity of endpoints in clinical trials. Any adjustment for multiple testing through adaptation of significance levels limits the statistical power of such analyses and may, therefore, not be the method of choice. Although this is true for statistical methods commonly used for analysis of QLQ-C30 data, such as linear mixed models, adjustment of significance levels is even more problematic when using a time-to-QOL deterioration method[34] or when conducting noninferiority trials.

The time-to-QOL deterioration method is a survival analysis approach that defines a deterioration in PRO scores of a given magnitude as an event. Although this approach is attractive because it can facilitate clinical interpretation of PRO scores, it may suffer from limited statistical power.[35] In noninferiority trials investigating whether one treatment is equivalent to another treatment, statistical power is also an issue, as these trials frequently target smaller effects than superiority trials. In both cases, the additional loss of statistical power makes correction for multiple testing an unattractive option. If in noninferiority trials the focus is mainly on noninferiority across all measured health domains, rather than on specific domains, the summary score may better reflect the research question being posed. Similarly, if the focus is on general health, further potential benefit of the QLQ-C30 Summary Score may be its use for assessing comparability of trial populations across studies or of study groups at baseline. Cancer registries collecting PRO data may also consider the QLQ-C30 Summary Score for comparing various patient groups. In addition, the QLQ-C30 Summary Score may be useful in the context of benchmarking, if heterogeneous patient groups are being compared.

Conclusion

In conclusion, the additional support for the single-factor measurement model for the QLQ-C30 Summary Score and its generally positive performance in generic group comparisons supports mainly its use in studies of patients with heterogeneous hematologic malignancies, particularly in those cases where symptom or other health domain-specific hypotheses are not available. The availability of QLQ-C30 Summary Score may help reduce potential problems with multiple testing that arise from using the individual QLQ-C30 questionnaire scales, and may facilitate the design of future clinical trials in patients with hematologic malignancies.

Acknowledgments

We would like to thank all the patients who participated in these Italian Group for Adult Hematologic Diseases (GIMEMA) studies and completed the quality of life questionnaires.

This analysis was funded by the Italian Group for Adult Hematologic Diseases (GIMEMA); GIMEMA Data Center.

REFERENCES

1. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst*. 1993;85(5):365–376.
2. Efficace F, Fayers P, Pusic A, et al. Quality of patient-reported outcome reporting across cancer randomized controlled trials according to the CONSORT patient-reported outcome extension: a pooled analysis of 557 trials. *Cancer*. 2015;121(18):3335–3342.
3. Burke JM, van der Jagt RH, Kahl BS, et al. Differences in quality of life between bendamustine-rituximab and R-CHOP/R-CVP in patients with previously untreated advanced indolent non-hodgkin lymphoma or mantle cell lymphoma. *Clin Lymphoma Myeloma Leuk*. 2016;16(4):182–190.e181.
4. Efficace F, Mandelli F, Avvisati G, et al. Randomized phase III trial of retinoic acid and arsenic trioxide versus retinoic acid and chemotherapy in patients with acute promyelocytic leukemia: health-related quality-of-life outcomes. *J Clin Oncol*. 2014;32(30):3406–3412.
5. Brundage M, Blazeby J, Revicki D, et al. Patient-reported outcomes in randomized clinical trials: development of ISOQOL reporting standards. *Qual Life Res*. 2013;22(6):1161–1175.
6. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA*. 2013;309(8):814–822.
7. Giesinger JM, Kieffer JM, Fayers PM, et al. Replication and validation of higher order models demonstrated that a summary score for the EORTC QLQ-C30 is robust. *J Clin Epidemiol*. 2016;69:79–88.
8. Gundy CM, Fayers PM, Groenvold M, et al. Comparing higher order models for the EORTC QLQ-C30. *Qual Life Res*. 2012;21(9):1607–1617.
9. Hinz A, Weis J, Faller H, et al. Quality of life in cancer patients—a comparison of inpatient, outpatient, and rehabilitation settings. *Support Care Cancer*. 2018;26(10):3533–3541.
10. Bower H, Bjorkholm M, Dickman PW, Hoglund M, Lambert PC, Andersson TM. Life expectancy of patients with chronic myeloid leukemia approaches the life expectancy of the general population. *J Clin Oncol*. 2016;34(24):2851–2857.
11. Lo-Coco F, Avvisati G, Vignetti M, et al. Retinoic acid and arsenic trioxide for acute promyelocytic leukemia. *N Engl J Med*. 2013;369(2):111–121.
12. Thanarajasingam G, Minasian LM, Baron F, et al. Beyond maximum grade: modernising the assessment and reporting of adverse events in haematological malignancies. *Lancet Haematol*. 2018;5(11):e563–e598.
13. Efficace F, Gaidano G, Lo-Coco F. Patient-reported outcomes in hematology: is it time to focus more on them in clinical trials and hematology practice? *Blood*. 2017;130(7):859–866.
14. Efficace F, Cannella L. The value of quality of life assessment in chronic myeloid leukemia patients receiving tyrosine kinase inhibitors. *Hematology Am Soc Hematol Educ Program*. 2016;2016(1):170–179.
15. Acquadro C, Regnault A. Patient-reported outcomes in drug development for hematology. *Hematology Am Soc Hematol Educ Program*. 2015;2015:496–500.
16. Venditti A, Piciocchi A, Candoni A, et al. Risk-adapted, MRD-directed therapy for young adults with newly diagnosed acute myeloid leukemia: results of the Aml1310 trial of the Gimema group. *Haematologica*. 2017;102(s2):6. abstract n. S111.
17. Castagnetti F, Breccia M, Gugliotta G, et al. Nilotinib 300 mg twice daily: an academic single-arm study of newly diagnosed chronic phase chronic myeloid leukemia patients. *Haematologica*. 2016;101(10):1200–1207.
18. Efficace F, Castagnetti F, Martino B, et al. Health-related quality of life in patients with chronic myeloid leukemia receiving first-line therapy with nilotinib. *Cancer*. 2018;124(10):2228–2237.
19. Efficace F, Baccarani M, Stagno F, et al. Health-related quality of life and symptom burden in chronic myeloid leukemia patients treated with first line with imatinib versus those treated with dasatinib: first results of the Leonidas study. *Blood*. 2017;130(suppl 1): 3422-3422.
20. Angelucci E, Santini V, Di Tucci AA, et al. Deferasirox for transfusion-dependent patients with myelodysplastic syndromes: safety, efficacy, and beyond (GIMEMA MDS0306 Trial). *Eur J Haematol*. 2014;92(6):527–536.
21. Efficace F, Santini V, La Nasa G, et al. Health-related quality of life in transfusion-dependent patients with myelodysplastic syndromes: a prospective study to assess the impact of iron chelation therapy. *BMJ Support Palliat Care*. 2016;6(1):80–88.
22. Efficace F, Gaidano G, Breccia M, et al. Prognostic value of self-reported fatigue on overall survival in patients with myelodysplastic syndromes: a multicentre, prospective, observational, cohort study. *Lancet Oncol*. 2015;16(15):1506–1514.

23. Efficace F, Niscola P, Patriarca A, et al. Pretreatment health-related quality of life profile according to the EORTC QLQ-C30 in patients with myelodysplastic syndromes (MDS): analysis on 443 lower-risk MDS patients. *Blood*. 2018;132(suppl 1):2293.
24. Hu L-t, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling*. 1999;6(1):1–55.
25. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. *Methods Psychol Res*. 2003;8(2):23–74.
26. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Rev ed. New York, NY: Academic Press; 1977.
27. Rosseel Y. lavaan. An R Package for Structural Equation Modeling. *J Stat Softw*. 2012;48(2):36.
28. Malcovati L, Germing U, Kuendgen A, et al. Time-dependent prognostic scoring system for predicting survival and leukemic evolution in myelodysplastic syndromes. *J Clin Oncol*. 2007;25(23):3503–3510.
29. Schumacher A, Wewers D, Heinecke A, et al. Fatigue as an important aspect of quality of life in patients with acute myeloid leukemia. *Leuk Res*. 2002;26(4):355–362.
30. Alibhai SM, Leach M, Kermalli H, et al. The impact of acute myeloid leukemia and its treatment on quality of life and functional status in older adults. *Crit Rev Oncol Hematol*. 2007;64(1):19–30.
31. Mols F, Husson O, Oudejans M, Vlooswijk C, Horevoorts N, van de Poll-Franse LV. Reference data of the EORTC QLQ-C30 questionnaire: five consecutive annual assessments of approximately 2000 representative Dutch men and women. *Acta Oncol*. 2018;57(10):1381–1391.
32. Hinz A, Einenkel J, Briest S, Stolzenburg JU, Papsdorf K, Singer S. Is it useful to calculate sum scores of the quality of life questionnaire EORTC QLQ-C30? *Eur J Cancer Care (Engl)*. 2012;21(5):677–683.
33. King MT, Bell ML, Costa D, Butow P, Oh B. The Quality of Life Questionnaire Core 30 (QLQ-C30) and Functional Assessment of Cancer-General (FACT-G) differ in responsiveness, relative efficiency, and therefore required sample size. *J Clin Epidemiol*. 2014;67(1):100–107.
34. Bonnetain F, Dahan L, Maillard E, et al. Time until definitive quality of life score deterioration as a means of longitudinal analysis for treatment trials in patients with metastatic pancreatic adenocarcinoma. *Eur J Cancer*. 2010;46(15):2753–2762.
35. Anota A, Barbieri A, Savina M, et al. Comparison of three longitudinal analysis models for the health-related quality of life in oncology: a simulation study. *Health Qual Life Outcomes*. 2014;12:192.