Original Research Article

MDM Policy & Practice

MDM Policy & Practice
2023, Vol. 8(2) 1–15

DOI: 10.1177/23814683231186992
journals.sagepub.com/home/mpp
Sage

Confirmatory Factor Analysis and Measurement Invariance of the Functional Assessment of Cancer Therapy Lung Cancer Utility Index (FACT-LUI)

J. Shannon Swan and Michelle M. Langer

Abstract

Background. A portion of the Functional Assessment of Cancer Therapy-Lung (FACT-L) instrument contributed to a previously published utility index, the FACT Lung Utility Index or FACT-LUI. Six FACT items representing lung cancer quality of life covered fatigue, pain, dyspnea, cough, anxiety, and depression. Two FACT items had been previously combined by the index authors into one for nausea and/or appetite loss, resulting in 7 final domains. **Methods.** The objective was to perform measurement invariance testing within a confirmatory factor analysis (CFA) framework to support the feasibility of using the FACT-LUI for non–preference-based psychometric applications. The original index patients comprised group 1, and similar FACT patient data (n = 249) from another published study comprised group 2. One 2-factor model and two 1-factor CFA models were evaluated to assess measurement invariance across groups, using varying degrees of item parceling and a small number of residual covariances, all justified by the literature. **Results.** The 1-factor models were most optimal. A 1-factor model with 1 pair of items parceled showed invariance to the partial scalar level using usual fit criteria across groups, requiring 2 unconstrained intercepts. A 1-factor model with 3 pairs of justified parcels showed full configural, metric, and scalar invariance across groups. **Conclusions.** The FACT-LUI items fit a partially to fully invariant 1-factor model, suggesting feasibility for non–preference-based applications. **Implications.** Results suggest useful incorporation of the FACT-LUI into clinical trials with no substantial increased respondent burden, allowing preference-based and other psychometric applications from the same index items.

Highlights

- This work suggests that in addition to being originally designed for use as a utility index, the 7 FACT-LUI items together also fit simple CFA and measurement invariance models. This less expected result indicates that these items as a group are also potentially useful in non–preference-based applications.
- Clinical trials can make for challenging decisions concerning which patient-reported outcome measures to include without being burdensome. However, the literature suggests a need for improved reporting of quality of life in lung cancer in particular as well as cancer in general. Inclusion of more disease-specific items such as the FACT-LUI may allow for information gathering of both preference-based and non–preference-based data with less demand on patients, similar to what has been done with some generic instruments.

Corresponding Author
J. Shannon Swan, Massachusetts General Hospital Institute for Technology Assessment, 101 Merrimac St 10th Floor, Boston, MA 02114, USA; (jsswan@mgh.harvard.edu).

Keywords
health-related quality of life measurement, psychometrics, factor analysis, structural equation modeling, utility index, preference-based measurement

Date received: September 19, 2022; accepted: June 8, 2023

Background

Health-related quality of life (HrQoL) measurement reporting in non–small-cell lung cancer has been reported as being of poor quality in the recent literature and must improve to support decision making.[1] Newer factors that influence lung cancer HrQoL are also becoming recognized that require closer monitoring by clinicians, including the effects of molecular alterations in neoplasms and newer treatments.[2,3] Beyond clinical trial reporting, HrQoL is also very important to patients. Recently, one study showed that in early lung cancer, HrQoL was more important than life extension, which is surprising given that life extension is usually the priority.[4] The need for better HrQoL reporting has also been called for in cancer trials in general.[5]

HrQoL reporting improvements should follow an efficient approach in which a substantial amount of information can be gathered in a small number of survey items. HrQoL data should ideally cover both measurement traditions, that is, the psychometric end of HrQoL as well as preference-weighted approaches. Concerning the latter tradition, lung cancer is one of the most costly neoplasms to treat,[6] and so cost-effectiveness is at issue as well. As noted by others,[7,8] it was not previously expected that indices obey usual psychometric stringencies given an assumed formative structure. Nonetheless, a result of the growing diversity in index development methods has been a greater cross-fertilization of quantitative traditions amid growing recognition of the ability of model domains to be judgmentally independent while being somewhat environmentally correlated.[9] The value of psychometrics has been embraced, not only for the initial development of domains and items but also for evaluation of their latent structure. Significant work in these developments include the National Health Measurement Study, which evaluated the latent variable structure of current generic indices.[10,11] Similarly, the Patient-Reported Outcomes Measurement Information System (PROMIS) initiative developed item banks with item response theory (IRT)–based quality assurance. The PROMIS Preference (PROPr) generic preference-based index was developed from PROMIS items and uses additional IRT-related methods.[12] These developments suggest an opening for the wider application of psychometric methods in index development, evaluation, and application. Consistent with this process, there have been recent structural equation modeling (SEM) techniques applied to the EQ-5D, an index composed of only 5 items.[13]

Reporting of cancer HrQoL from the psychometric end of the field has previously used the European Organization for Research and Treatment of Cancer (EORTC) surveys[14] and the Functional Assessment of Cancer Therapy (FACT) instruments,[15] as well as generic instruments including the Short Form (SF)-12 and SF-36[16,17] among others. Methods used for construction of such measures rely on classical test theory, factor analytic models, and, more recently, IRT.[12,18] These approaches have been critical for accurate documentation of patient-reported outcomes, addressing psychometrics at the scale and item level. For the preference-based end of HrQoL in cancer, most efforts have been more generic in scope, using the usual societal indexes (EQ-5D, HUI2/3, SF6D[19]) or adaptations covering cancer in general using EORTC or FACT[20,21] with a small amount of prior work addressing lung cancer specifically.[22,23]

As a potential contributor to a solution for lung cancer–specific HrQoL-reporting problems indicated above, a recently published preference-based index for lung cancer HrQoL, the FACT Lung Cancer Index, or FACT-LUI, was developed from a very parsimonious subset of items from the FACT-L (Lung) instrument for lung cancer patient-reported outcomes.[23] Swan et al.[23] used 6 of the items in their original form, and 1 item is a

Massachusetts General Hospital Institute for Technology Assessment, Boston, MA, USA (JSS); Harvard Medical School, Boston, Massachusetts, USA (JSS); Northwestern University, Feinburg School of Medicine, Chicago, Illinois, USA (ML). The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: American Cancer Society (#126904-PEP-14-206-01-PCSM). Financial support for this study was provided entirely by a grant from the American Cancer Society. The funding arrangement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

combination of 2 original items as described below. While the FACT-LUI initially appears promising from the standpoint of preference-based measures, it is derived from a small number of items from a larger source of FACT items. Therefore, one cannot assume that the measurement properties of the FACT-LUI are adequate from a psychometric perspective. In other words, one must show that the item group still "measures what it is intended to measure" from this point of view. A reasonable approach to such a problem is to assess measurement invariance within a confirmatory factor analysis (CFA) model for the FACT-LUI, including assessments of global and local fit. We elected to evaluate measurement invariance over the breadth of lung cancer morbidity by comparing FACT-LUI patient data from Swan et al.[23] to similar FACT-L data from a second patient group undergoing therapy.[24] This difference in morbidity between the groups provides a convenient test of invariance to which a measure should be robust if it is of sound construction.

For guidelines in reporting, 2 sources were used. We used table 4.7 on reporting results from a recent widely used text on CFA.[25] For our invariance table and figures, we adapted the recommendations of Putnick and Bornstein.[26]

Methods

Ethics Approval

The questionnaire and methodology for this study were approved by the institutional Human Research Ethics committee at our hospital (Mass General Brigham IRB at Massachusetts General Hospital, protocol: 2014P002045). As directed by our ethics committee, consent was implied by completing the surveys, and the committee directed us on the composition of the recruitment letter. This letter was also designed to incorporate the elements of informed consent.

Domains of the FACT-LUI

The 7 domains of the FACT-LUI originated with a review of lung cancer HrQoL research described in Swan et al.,[23] particularly using the symptom cluster literature, with the intent of capturing the aspects that were most predictive of lung cancer HrQoL.[27–30] Furthermore, given the usual structure of preference-based indices, a parsimonious group of domains was chosen by the authors that could be judgmentally independent for valuation while recognizing the possibility of correlated symptoms. The FACT-LUI domains that are represented by 1 FACT-L item using its original language included fatigue (I have a lack of energy), pain (I have pain), anxiety (I worry that my condition will get worse), depression (I feel sad), cough (I have been coughing), and dyspnea (I have been short of breath). For the remaining domain, the index authors had concerns for judgmental dependence that were great enough to combine the original separate item content for nausea and appetite, thus a new item: "I have nausea and/or appetite loss." This item combination approach for nausea and appetite issues was justified by literature support and has reportedly been used by others in preference-based index development.[31,32] Each FACT-LUI item has a 5-point response set for degree of severity identical to the FACT-L and other FACT lung cancer–related instruments mentioned below, that is, none, a little bit, somewhat, quite a bit, very much. The initial published index results showed good construct validity when comparing advanced versus early-stage patients and by agreement with directly elicited utilities.[23]

Patient Data (Table 1)

*First data set for invariance testing (*n *= 237): group 1.* In data designated for group 1 in our prior work,[23] Table 1 patients could be at any stage of cancer and any point in their disease trajectory, from diagnosis to follow-up. Patients with surveys in hand completed interviews by telephone, facilitated by a trained interviewer with the permission of the attending oncologist. All survey instruments had local institutional review board approval.

*Second data set for invariance testing (*n *= 249): group 2.* Lung cancer patient data were provided from a study by Yount et al.[24] (Table 1) that included FACT item content similar to the FACT-LUI. As we indicate below, there are multiple FACT instruments (1-wk recall), and some items as used in the FACT-G are, as the name abbreviation suggests, more generic in scope for most cancers, while others are more disease specific, for example, FACT-L (Lung), FLSI (FACT Lung Cancer Symptom Index), and FACT-LCS (Lung Cancer Subscale), among others. In the Yount et al. study, patients completed surveys in a mixed-mode fashion (touch screen tablets, interview, and interactive voice-response technology) during a 12-wk study in which all were in treatment. We chose the Yount et al. first survey assessment, since 2 FACT instruments that together contain the content for the FACT-LUI items (FACT-G[15] and FACT-LCS[33]) were completed at that time. Five of the 8 FACT-L items used in the FACT-LUI item group

Table 1 Demographics of Patient Groups

	Group 1: from Swan et al. (2018)[23,a]	Group 2: from Yount et al. (2014)[24,a]
Sample size	237	253[b]
Age, y, $\bar{x}$ (*s*)	65.4 (10.4)	60.6 (10.2)
Gender, male, *n* (%)	101 (42.6)	125 (49.4)
Race/ethnicity, *n* (%)		(missing = 1)/(missing = 4)
White	217 (91.6)	147 (58.3)
Black	5 (2.1)	91 (36.1)
Other	15 (6.3)	14 (5.6)
Hispanic	5 (2.1)	11 (4.4)
Education		(missing = 3)
12 y or less, *n* (%)	58 (24.5)	
Years, median (IQR)	16 (13,18)	
Eighth grade or less		9 (3.6)
Some high school		36 (14.4)
High school graduate/GED		73 (29.2)
Some college/tech/AA		59 (23.6)
College degree (BA/BS)		41 (16.4)
Advanced degree (MA, PhD, MD)		32 (12.8)
Diagnosis		(missing = 17)
NSCLC	237	204
SCLC		32
Stage, *n* (%)		(missing = 11)
I-II	73 (30.8)	
III-IV	164 (69.2)	
IIIa		31 (12.8)
IIIb		53 (21.9)
IV		134 (55.4)
SCLC		24 (9.9)

IQR, interquartile range; NSCLC, non–small-cell lung carcinoma; SCLC, small-cell lung carcinoma.
[a]Original data sources, see text.
[b]Usable sample for this study was $n = 249$; see missing data discussion in text.

are found in the FACT-G (pain, fatigue, nausea, depression, anxiety) and 3 (dyspnea, cough, appetite) are also present in the FACT-LCS survey. The dyspnea, fatigue, pain, and cough items also comprise 4 of the items in the FLSI, a 6-item version of the FLSI-12.[34] The original FACT items mentioned here all have identical item language in these FACT instruments. Because the FACT-LUI items are a subset, they are not referred to as a FACT instrument here.

Model Specification (Figure 1)

Justification of 1- and 2-factor models and use of parceling. Due to the intended parsimonious domain/item set of the FACT-LUI, the domains were initially envisioned as grouped into 2 factors, based on the lung cancer symptom cluster literature. A study of 2,405 patients by Cheville et al.[35] showed an enduring cluster of fatigue, dyspnea, and cough. Given the Cheville et al. study's substantial sample size and methods used, having these domains in a factor together was a priority, relegating the other domains of pain, anxiety, depression, and nausea and/or appetite loss to a second factor. A 400-person study by Henoch et al.[28] proposed a respiratory cluster of breathing and cough as well as a mood cluster. Furthermore, fatigue and appetite had high loadings on pain (which included gastrointestinal issues) and respiratory items. Other work by Carnio et al.[36] reviewing the effects of fatigue suggested a contribution of fatigue to the pain and psychological dimensions. Given such overlap in domains and other variations in clusters in the literature,[29] it was unsurprising that we found a strong correlation between the 2 proposed factors ($r = 0.74$–0.87 for groups 2 and 1, respectively, in initial analysis). Due to these results, we report the 2-factor model, but a 1-factor structure was also modeled going forward (Figure 1).

The FACT-LUI combined nausea and/or appetite loss item as used in group 1 presented a situation in which

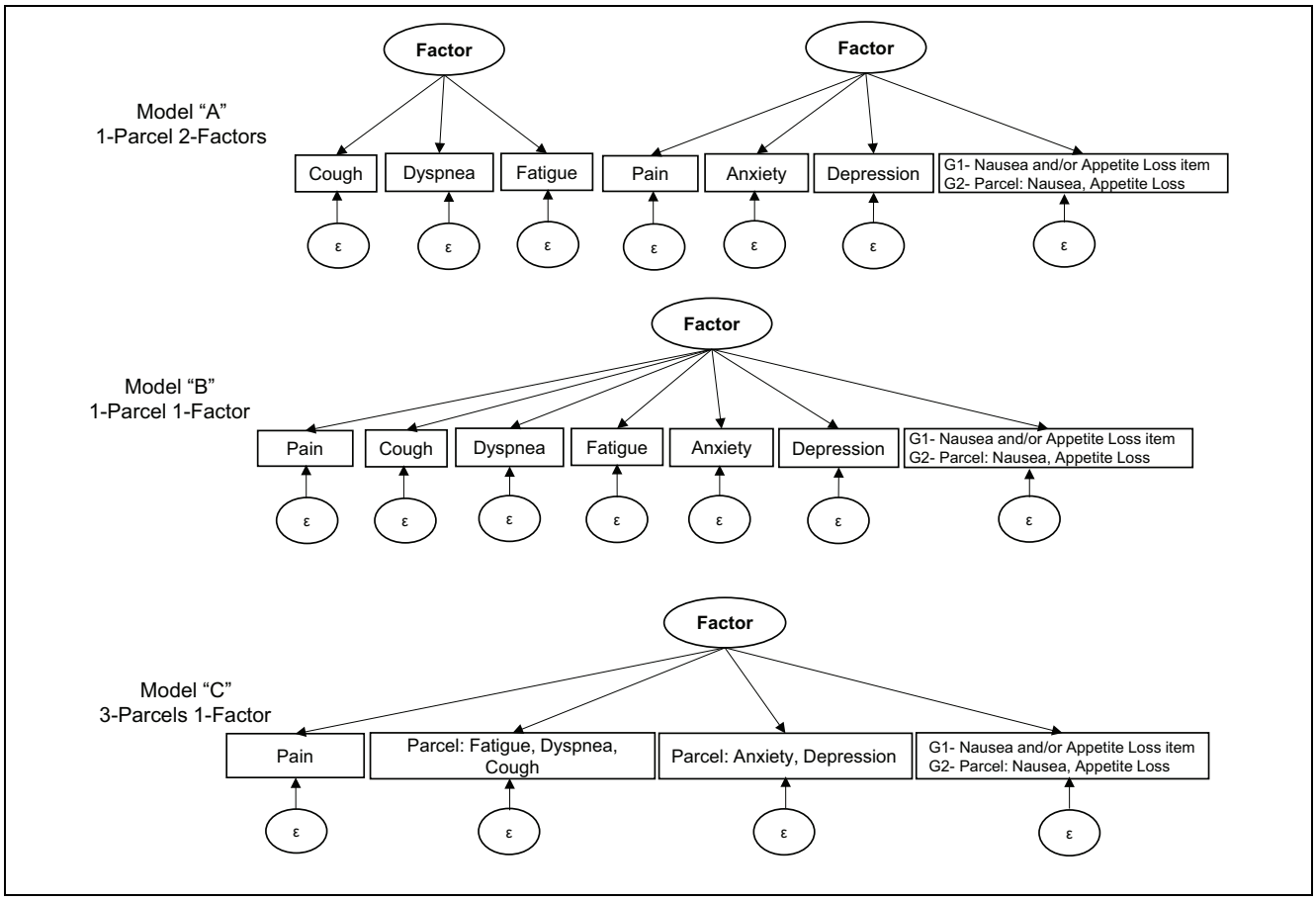

Figure 1 Model Specifications - Model "A" has 2-factors and 1-parcel; Model "B" has 1-parcel and 1-factor, and Model "C" has 3 parcels and 1-factor for CFA of the FACT-LUI index items.
G1 refers to Group 1 where the original FACT-LUI nausea and/or appetite loss item was assumed to act as a parcel. G2 refers to Group 2 where original separate FACT items for nausea and appetite were parceled. The symbol ε refers to item residual variance. The 2-Factor model is based on literature as described in text.

invariance testing of group 1 against the group 2 data set was not initially possible, since nausea and appetite loss were represented by the 2 original and separate FACT items in group 2 ("I have a good appetite" and "I have nausea"). We rectified this situation by investigating the following approach. It was our hypothesis that the "and/or" item would be equivalent to an average parcel of the 2 original FACT items, scored similarly for best and worst levels. Item parcels are often used to minimize item-specific variance in CFA while preserving common construct variance, usually by averaging 2 or more items to create a new indicator.[37] To provide evidence for our assertion, we used provided data obtained in group 1 patients using the EORTC QLQ-30 v.3 survey.[14] The EORTC nausea item ("Have you felt nauseated?") and the appetite item ("Have you lacked appetite?") use language for the response set identical to the FACT survey, except that there are 4 options (not at all, a little, quite a bit, very much) instead of the 5 used in FACT instruments, thus no "somewhat" option. The EORTC and FACT data from the group 1 patients had been obtained by the index authors during the same interview.

Using the meta-analysis and SEM literature as a guide,[38,39] we linearly transformed the group 1 EORTC nausea and appetite items to a 5-point scale. Both items were transferred to the 5-point FACT scale, which ranges from 0 to 4, where 0 is the worst value and 4 as the best, as appropriate for each item's content. We converted the 2 transformed 5-point EORTC-derived items to an average parcel and compared them for agreement with the original 5-point "and/or" item in group 1. Using a mixed intraclass correlation coefficient (ICC) model in which

the same raters were used and assuming strict agreement, an ICC of 0.87 for average ratings (0.77 single measures) was found between the EORTC parcel and the original "and/or" item. ICCs for single and average ratings are reported since one comparator is based on averages from parceling. Mountain Plots[40] showed a 0.00 median bias between the EORTC parcel and the "and/or" item, with a mean difference of 0.1 on the 0 to 4 FACT scale (10% trimmed mean 0.07) and 0.25 mean absolute difference (10% trimmed mean 0.19) between comparators. Given this evidence that the original "and/or" item was a reasonable parcel approximation in group 1, it was applied to our 1-factor model specifications for nausea and appetite loss (Figure 1). Considering the symptom cluster literature[27–29,41] and clinical sources, further parcels were later formed for another model by matching domain content, based on the aim that a parcel may provide more complete representation of domain content or what has been called the construct centroid[37] in a parcel-based indicator. We felt strongly that parcels should be clinically justified and not randomly constructed.

In our models A and B (Figure 1), the a priori structure of the original index was replicated, and parceling was used sparingly given its sometimes controversial status.[37] Thus, the original FACT-LUI with 7 items/domains (the original nausea and/or appetite item along with the other 6 original FACT items from group 1) was used against an average parcel of the 2 original group 2 nausea and appetite items only, while the other group 2 items were left as individual indicators. As above, these 2 models differ by having 2 factors in model A and 1 factor in model B.

The third model was a 3-parcel version of the 1-factor specification (Figure 1), where, in addition to parceling the original nausea and appetite items from group 2 against the "and/or" item in group 1, 2 other parcels were created in each group. First, anxiety and depression, each represented by 1 item, are widely known to clinically co-occur in most patients. These diagnoses are often argued as being on the same clinical continuum[42–48]; thus, these 2 items became a parcel. As a further example of this construct, the noted PHQ-4 instrument summates 2 anxiety and 2 depression items for an emotional distress score.[49] Second, as above, the Cheville et al.[35] study justified our placing fatigue, dyspnea, and cough in a parcel given their symptom cluster relationship.

All items were analyzed using the proper item score meaning on the FACT ordinal 0 to 4 scale where 0 = the worst level and 4 = best level prior to being parceled or left as separate. Pain was left separate due to its less clear association with other symptom clusters.[28,29,50]

Estimator and Measures of Fit

Initial evaluation showed the expected skewed/kurtotic distributions in the FACT-LUI items and a lack of multivariate normality (Doornik-Hansen and Mardia estimates of $P < 0.001$). Given that the FACT items all have 5-point Likert response sets, the robust version of maximum likelihood was justified as an estimator as noted by others.[51–54]

For CFA modeling, MPlus v8.5 (Muthén and Muthén, Los Angeles, CA, 2017) was used. To evaluate other comparisons and demographics, MedCalc v.19.2.1 (Medcalc Software Ltd, Ostend, Belgium) and Stata v.16.1 (StataCorp. 2019. College Station, TX: StataCorp LLC) were used. To evaluate model fit, we reported χ^2 statistics, with robust correction[54] for the nonnormality of data by the Satorra-Bentler method[55] with usual P values greater than or equal to 0.05 preferred. Measures of model fit[56–58] included the root mean square error of approximation, where point of estimates less than 0.06 were preferred and 0.08 preferred at the upper limits of the 90% confidence interval (CI). The Comparative Fit Index was used, with values at 0.95 or greater a guideline for good fit. We also obtained the standardized root mean square residual with values less than 0.08 preferred. For local fit, we focused on the residuals for correlations, with values greater than 0.10 being less optimal. We evaluated relative morbidity between groups by item with independent group nonparametric tests.

For configural models in invariance testing, factor loadings and intercepts were free (unconstrained) across groups and factor means fixed at zero (for model identification) in all groups. The scale of a factor was set by freeing all factor loadings and fixing the factor variance to 1 in both groups. For the metric level of invariance testing, factor loadings were constrained to be equal across groups, with intercepts free across groups, and factor means fixed at zero across groups for model identification. The factor variance was fixed at 1 in group 1 and was free in the other group. At the scalar level, factor loadings and intercepts were constrained to be equal across groups with the factor mean fixed at zero in group 1 but free in group 2. The scale of the factor was set by fixing the factor variance to 1 in group 1 and free in group 2. If noninvariance was noted by modification indices (MIs), a parameter constraint was released and the model reestimated. Residual covariances

Table 2 Severity of Disease By FACT Items in Groups[a]

	Group 1				Group 2				
Variable	***n***	**Median**	**Mean**	**Mean Rank**	***n***	**Median**	**Mean**	**Mean Rank**	**Two-Tailed *P* Value**
Anxiety	237	3.00	2.65	245.44	245	3.00	2.59	237.69	0.53
Cough	237	3.00	3.17	276.95	246	3.00	2.57	208.33	<0.001
Depression	237	4.00	3.28	270.63	245	3.00	2.84	213.32	<0.001
Fatigue	237	3.00	2.54	263.61	249	2.00	2.17	224.36	0.002
Nausea-appetite loss[b]	236	4.00	3.55	291.58	243	3.00	2.96	189.91	<0.001
Pain	237	4.00	3.19	266.97	244	3.00	2.72	215.78	<0.001
Dyspnea	237	3.00	3.06	270.32	246	3.00	2.55	214.72	<0.001

FACT, Functional Assessment of Cancer Therapy.
[a]Mann-Whitney test comparison of items across groups.
[b]Item data are from the 1-parcel model, using the original nausea and/or item in group 1 and the average parcel for nausea and appetite items in group 2. Items are all on a 0 (worst) to 4 (best) ordinal scale as used in the FACT instruments.

were allowed in a model specification as suggested by MIs if supported by the literature.

External Correlates

As an assessment of the FACT-LUI items against an existing measure, a correlation matrix was constructed analyzing the correlation of the responses for the EORTC items available from group 1 with group 1 FACT-LUI item responses for items with similar concepts. Summated FACT-LUI items were also correlated with obtained visual analog scale measures from the prior study.[23]

Results

Demographics (Table 1)

The groups were relatively similar in size, (Table 1) with minor differences in age and gender demographics. Group 2 from Yount et al.[24] showed greater minority representation. There was a greater degree of formal education in group 1. Most patients had non–small-cell carcinoma with a minority in group 2 having small-cell carcinoma or no specific lung cancer cell type documented. FACT survey items are the same across all lung cancer cell types. Most patients had advanced disease, which is typical for the presentation of lung cancer.[59]

Data Distribution: Severity of Disease (Table 2)

Group 2 patients (Table 2) were expected to show more morbidity since they were all undergoing chemotherapy and all were in more advanced cancer stages. This expectation was confirmed by significant differences in all items but anxiety (Bonferroni adjustment at 0.007 for 7 items).

Missing Data (Table 2)

Group 1 had 1 missing entry for the nausea/appetite item with otherwise complete data for all 237 patients (Table 2). Four group 2 patients had no data for both the FACT-G and FACT-LCS instruments. Exclusion of the 4 group 2 patients with no data left 249 for further analysis. There were 4 group 2 instances of missing data for the nausea, depression, and anxiety items and 5 instances of missing data for the pain item. Three group 2 missing data points were seen for each of the dyspnea, cough, and appetite items. Any parceled indicators with missing item data for 1 or more items were designated as missing for that parceled indicator in a patient. These missing data resulted in a range of 243 to 249 group 2 patients with complete data by item for the 1-parcel model and 236 to 237 group 1 patients (Table 2). In the 3-parcel model, there was a range of 236 to 237 group 1 patients and a range of 243 to 245 group 2 patients with complete item data (not shown in Table 2). Full information maximum likelihood was used by default in MPlus.

Invariance Testing: 1-Parcel Model with 1- and 2-Factor Specifications (Figures 2 and 3, Table 3)

MIs for allowed residual item covariances (3.84 and higher considered) were used sparingly to reflect the known close relationship between residuals of some domains as noted in the above justifications (Figures 2

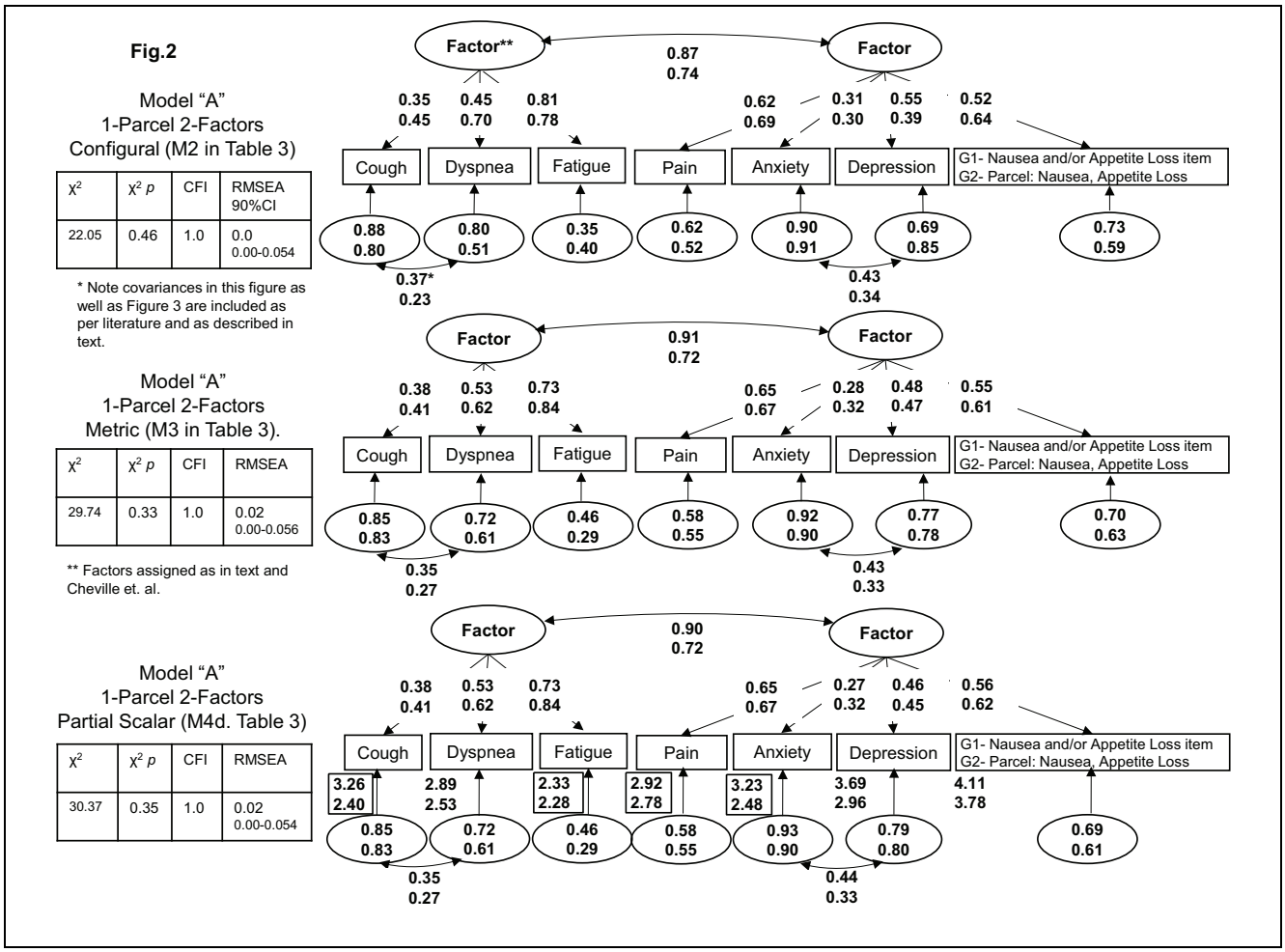

Figure 2. Measurement invariance parameter estimates for 1-parcel 2-factor model (Model "A").
Standardized estimates are shown for interpretability. Any inequalities are due to the factor and variable variances which are used for standardization and are being allowed to be unequal while unstandardized loadings are constrained to equality in the model. Residual variances (ellipses) are seen below each variable. 2-headed arrows and estimates show allowed residual covariances (see text). Each pair of parameter estimates show Group 1 results atop Group 2. Non-invariant (unconstrained freely estimated) intercepts according to Satorra-Bentler adjusted χ^2 difference tests are reported in a box below the affected variable. Abbreviated fit statistics are shown with each invariance testing stage. Complete fit details are shown in Table 3.

and 3, Table 3). Allowed residual covariances were added 1 at a time across groups followed by reestimation. As expected, the largest MI initially in the configural model was from the covariance between anxiety and depression residuals (MI 34.96 and 27.30 within groups 1 and 2, respectively, in the 1-factor model and 33.88 and 26.84, respectively, in the 2-factor model). Subsequently, the known symptom cluster of fatigue, cough, and dyspnea was included by MI results for dyspnea and cough and dyspnea and fatigue in the 1-factor model and dyspnea with cough in the 2-factor model. These covariances were included in both groups for the configural model and carried through the metric and scalar models. There were other residual covariances that could have been included with justification, but we felt they were not at the evidence level of the Cheville et al.[35] study and the literature on anxiety and depression. Following these meaningful respecifications, the metric versus configural comparison showed a nonsignificant χ^2 difference test in both models (Table 3). At the scalar level in the 1-factor model, the fatigue intercept constraint was released in the first step (MI 8.77 in both groups) and the pain intercept in the second (MI 5.7 in both groups after fatigue release), reflecting differences in the mean amounts of those

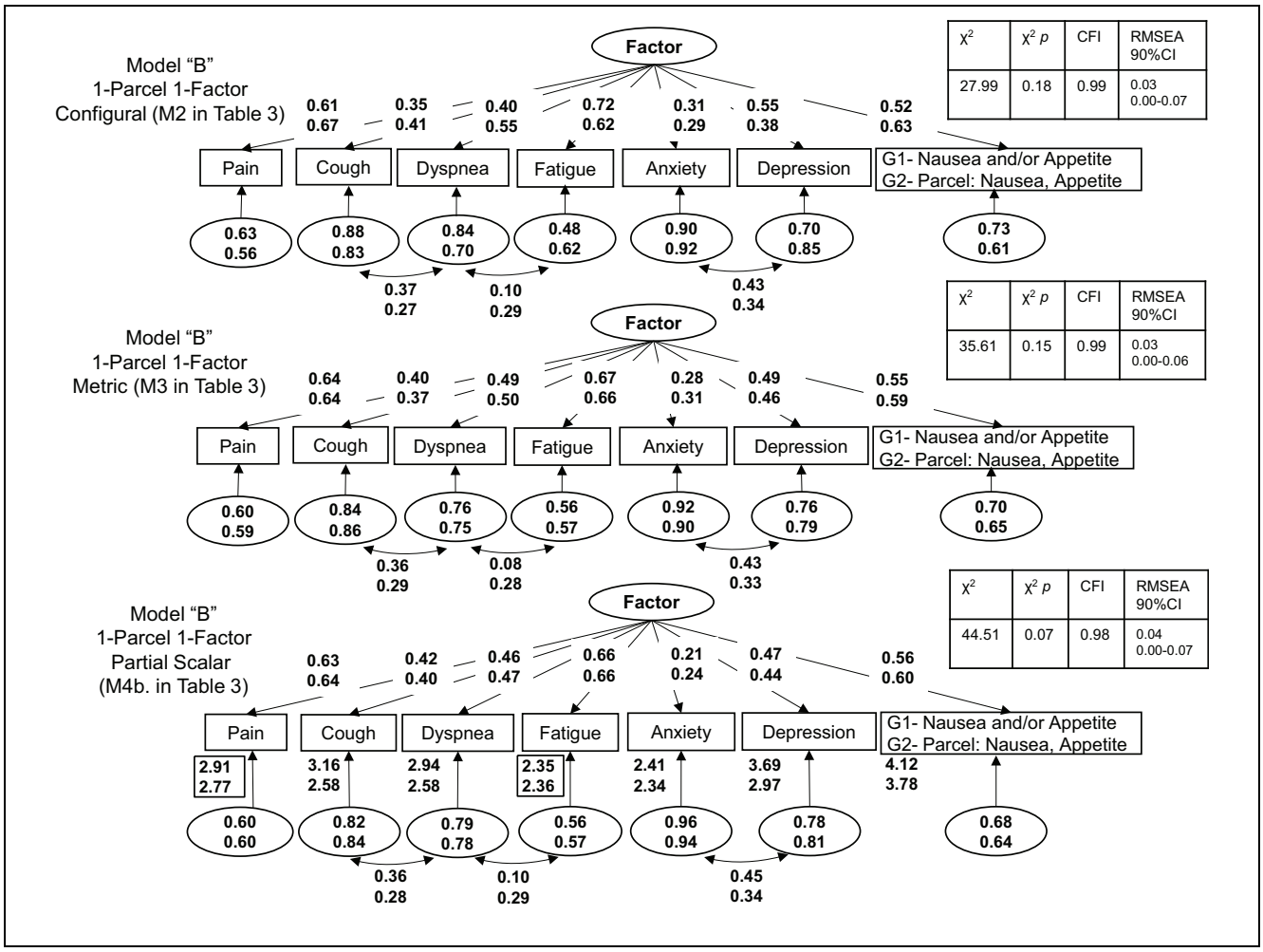

Figure 3. Measurement invariance parameter estimates for 1-parcel 1-factor model (Model "B").
Abbreviated fit statistics (see also Table 3), standardized parameter estimates and allowed residual covariances as described in text are shown.

specific variables across groups 1 and 2 above and beyond any differences in the underlying HrQoL construct. Following these respecifications, the (partial) scalar versus metric comparison yielded a statistically nonsignificant χ^2 difference in the 1-factor model. Remarkably, 4 intercepts required release in the 2-factor model to get to a nonsignificant χ^2 difference test. Other fit indices showed acceptable results at all invariance testing levels in both models (Table 3). Standardized parameter estimates are shown in Figures 2 and 3 and unstandardized output in supplemental data. There was little issue noted in local fit, with correlation residuals usually below 0.10 (Table 3).

Invariance Testing: 3-Parcel Model (Figure 4, Table 3)

The second 3-parcel model showed reasonable fit at the configural, metric, and scalar levels without further modification. χ^2 difference tests were statistically nonsignificant when comparing the configural to metric and the metric to scalar models (Figure 4, Table 3).

External Correlates

A correlation matrix constructed with prior EORTC data from group 1 was set against similar item domains

Table 3 Invariance Testing[a]

Standardized Results	χ^2 (*df*)	χ^2 (*P*)	Mean G1/G2	CFI	RMSEA (90% CI)	Prob.RMSEA <0.05	SRMR	Model Comp	$\Delta\chi^2$ (Δdf)	Corr resid	Decision
1-Parcel 2-factor											
Model A											
M2 configural	22.05 (22)	0.46	0.00/0.00	1.0	0.00 (0–0.05)	0.93	0.03	—	—	0/0	
M3 metric	29.74 (27)	0.33	0.00/0.00	1.0	0.02 (0–0.06)	0.91	0.04	M2	7.50^(5)	1/1	Accept
M4 scalar	56.34 (32)	0.01	0.00,0.00/−0.66−0.96	0.95	0.06 (0.03–0.08)	0.32	0.07	M3		1/1	
M4a. partial scalar	46.52 (31)	0.04	0.00,0.00/−1.07−0.96	0.97	0.05 (0.01–0.08)	0.59	0.06	M3		2/2	
M4b. partial scalar	41.11 (30)	0.09	0.00, 0.00/−0.8, −0.98	0.98	0.04 (0–0.07)	0.72	0.05	M3	12.14^^(3)	2/1	
M4c. partial scalar	35.70 (29)	0.18	0.00, 0.00/−1.06, −1.19	0.99	0.03 (0–0.06)	0.83	0.05	M3	6.39^^^(2)	1/1	
M4d. partial scalar	30.37 (28)	0.35	0.00, 0.00/−0.90, −1.19	1.0	0.02 (0–0.05)	0.92	0.05	M3	0.58^^^^(1)	1/1	Accept
1-Parcel 1-factor											
Model B											
M2 configural	27.99 (22)	0.18	0.00/0.00	0.99	0.03 (0−0.07)	0.76	0.03	—	—	0/0	—
M3 metric	35.61 (28)	0.15	0.00/0.00	0.99	0.03 (0–0.06)	0.80	0.05	M2	7.62 (6)*	1/1	Accept
M4 scalar	60.35 (34)	0.00	0.00/−0.76	0.95	0.06 (0.03–0.08)	0.30	0.07	M3		2/0	
M4a partial scalar	50.31 (33)	0.03	0.00/−0.88	0.97	0.05 (0.02–0.07)	0.56	0.06	M3		2/1	
M4b. partial scalar	44.51 (32)	0.07	0.00/−1.01	0.98	0.04 (0.00–0.07)	0.70	0.06	M3	9.0(4)**	3/1	Accept
3-Parcel 1-factor											
Model C											
M2 configural	1.54 (4)	0.82	0.0/0.0	1.0	0.0 (0.0–0.06)	0.93	0.01	—	—	0/0	—
M3 metric	4.40 (7)	0.73	0.0/0.0	1.0	0.0 (0.0–0.06)	0.92	0.04	M2	2.72***(3)	0/0	Accept
M4 scalar	11.12 (10)	0.35	0.0/−0.77	1.0	0.02 (0.0–0.08)	0.76	0.06	M3	6.59****(3)	0/0	Accept

χ^2, chi-square; CFI, comparative fit index; CI, confidence interval; Corr resid, correlation residual (number residuals $> \pm 0.10$ in G1/G2); (*df*), degrees of freedom; G1/G2, group 1/group 2; model comp, model comparisons for χ^2 difference test; $\Delta\chi^2$ (Δdf), chi-square difference test (Satorra-Bentler correction) and difference in degrees of freedom; (*P*), 2-tailed *P* value; prob., probability; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual.

[a]Factor model invariance steps are indicated by "M2" thru "M4" with additional as follows: M4a, release fatigue intercept; M4b, release pain intercept; M4c, release anxiety intercept; M4d, release cough intercept. Invariance modeling conventions: configural, no model restrictions; metric, loadings constrained across groups; scalar, all intercepts constrained (except as noted in text and figures and as "partial" scalar here) and loadings constrained across groups.

*$P = 0.27$; **$P = 0.061$; ***$P = 0.44$; ****$P = 0.086$.

^$P = 0.19$; ^^$P = 0.007$; ^^^$P = 0.04$; ^^^^$P = 0.45$.

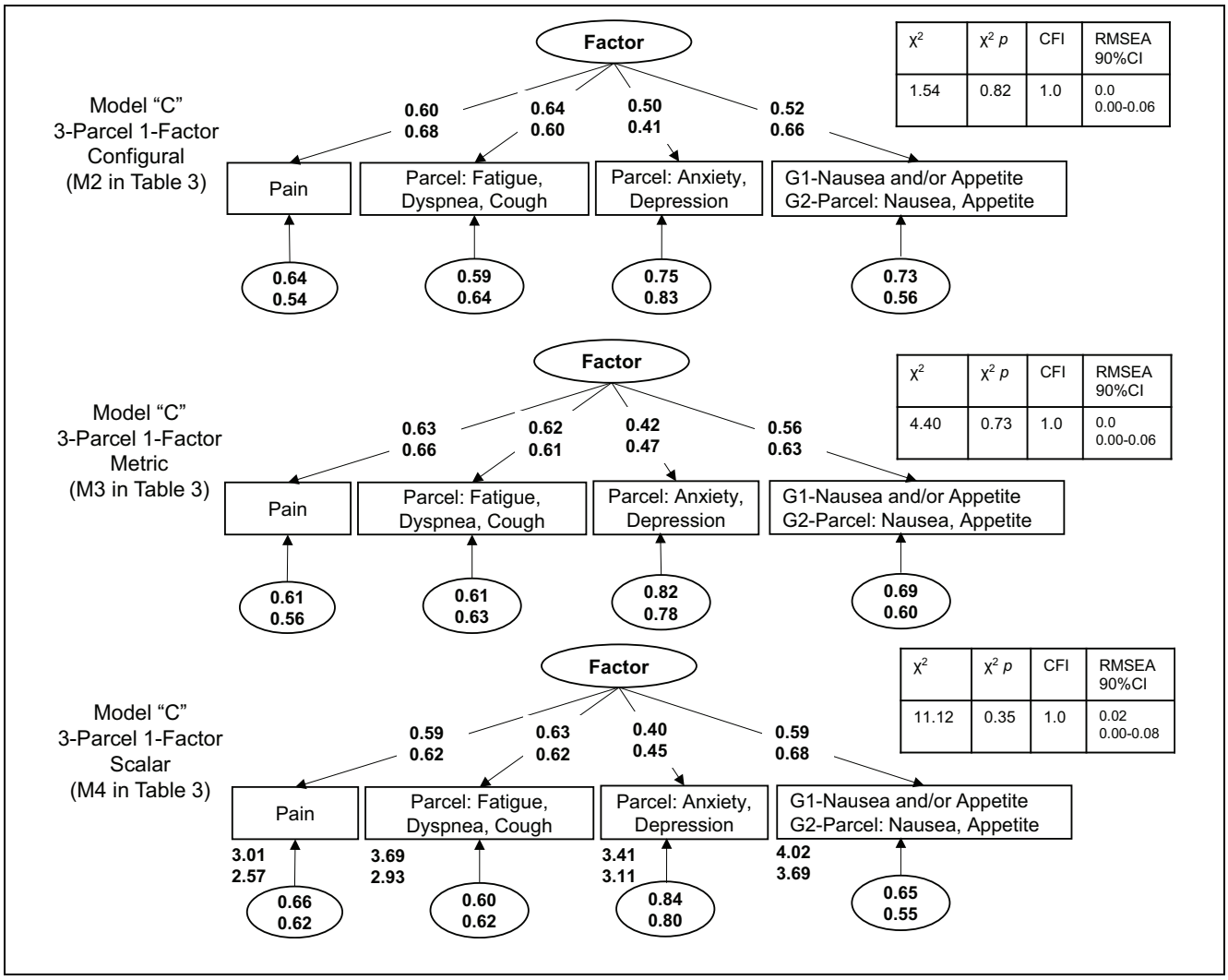

Figure 4 Measurement invariance parameter estimates for 3-parcel 1-factor model (Model "C").
Abbreviated fit statistics (See also Table 3) and standardized parameter estimates are shown.

from the FACT-LUI (see supplemental data). These data showed reasonably strong correlations with each other: depression items: $r = 0.78$, fatigue: $r = 0.72$, dyspnea: $r = 0.78$, pain: $r = 0.66$, cough: $r = 0.87$, nausea-appetite loss: $r = 0.80$, anxiety: $r = 0.65$. The summated scores of the EORTC items showed a 0.87 Pearson correlation with summated FACT-LUI items. Finally, in Swan et al.,[23] a correlation was shown between visual analog scale data in each patient and the sum of their FACT-LUI items of −0.60. This is expected because higher utility would be seen with less morbidity and thus a lower summated score for the items.

Discussion

The straightforward 1-factor model specification seemed clear from the results and was theoretically attractive since the original index is meant as a summary measure of lung cancer HrQoL. Although the FACT-LUI was developed with more of a formative frame of reference, CFA was still feasible in our case due to the ability of such constructs to fit reflective models while having adequate judgmental/structural independence during the index valuation process. Such independence is usually assumed by clinical domain selection and a lack of difficulty seen in respondents during index valuation sessions.

However, a more formal demonstration of feasibility of structural independence in an index was reported during development of the PROPr.[12,60]

Not all FACT-LUI items had favorable factor loadings, as was the case with the anxiety domain and to a lesser degree the cough domain. It could be argued that an item querying about worry that one's disease may progress might not adequately sample the anxiety morbidity present. In group 2, there were data for other FACT items available. To explore this finding, we evaluated an anxiety item of interest from FACT-G not in the FACT-LUI ("I feel nervous"). We suspected this item may be more encompassing of this construct. Interestingly, this item tapped into less morbidity than the original item did ($P = 0.02$, Wilcoxon signed rank, $n = 245$) in group 2. The 3-parcel model may have supported the apparent weaker loadings in anxiety and cough to better represent the construct centroid of these constructs via depression and dyspnea[37] items, respectively.

A criticism of parceling is that while it may minimize item-specific or unique variance and enhance communality, it could conceivably enhance model fit while obscuring multidimensionality.[25,37] Multidimensionality appeared unlikely in our models, given their small size and 2 highly correlated factors in initial work. Furthermore, the FACT-LUI being disease specific may have increased the likelihood of unidimensionality. In these 1-factor models, we observed full measurement invariance through the scalar level with a 3-parcel solution based on χ^2 difference testing, but we observed metric invariance in both the 1-parcel and 3-parcel models, requiring only 2 scalar intercept releases in the 1-parcel model. Leaving aside the χ^2 difference tests, all other measures gave acceptable results at the scalar level with no intercept releases in the 1-parcel model (Table 3). The proposed 3-parcel CFA model seems reasonable in its context based on the clinical co-occurrence of anxiety with depression, nausea with appetite loss, and the chronic symptom cluster of fatigue, cough, and dyspnea.

Invariance results were encouraging given an expected degree of group morbidity differences that could stress a model (Table 2). The only domain without a significant difference between groups was anxiety, although its mean item score still reflected greater morbidity in group 2. This result may be explainable on the basis that lung cancer is a devastating disease; hence, anxiety about disease progression may be considerable, no matter where in the disease trajectory one is located.

Due to the usual broad conceptualization of preference-based indices, it might be argued that some substantial aspect of this CFA could be formative in its modeling. Recent work with the EQ-5D has investigated specifications that variously incorporated items as formative.[13] However, the distinction of formative versus reflective items is much clearer cut in the EQ-5D than in the FACT-LUI. As described above, the latter has items of a greater symptom orientation while still covering the pain, physical, and psychological domains needed in utility indices.[11]

CFA studies are often to some degree exploratory since some respecification is generally needed. Our use of modification indices for a small number of residual covariances was justified by the work of others. Fortunately, there is a large literature on quality of life in cancer, particularly for symptom clusters that help to inform CFA models of this type.

In general, other model specifications such as second-order and bifactor approaches might be considered. In our case, the 2-factor model (model A) would be the only alternative in such specifications. A second-order model is infeasible because a second-order factor model with 2 first-order factors as indicators is underidentified. Our application of a bifactor approach added to the 2-factor model caused multiple Heywood cases. When these cases are resolved by constraining residual variances in the affected items, a bifactor structure did not contribute to the model substantively, since the 2 residualized factors used from model A had little variance left to explain outside of the general factor. This was indicated by multiple nonsignificant residualized factor loadings and by a high correlation of nonstandardized loadings in the prior 1-factor model with the general factor of the bifactor model ($r = 0.98$, 95% CI 0.9486–0.9951). Therefore, a 1-factor model is strongly suggested.

Limitations

In addition to usual concerns about the existence of multidimensionality being obscured by parceling, it is often done in a more random fashion,[61] which calls into question the validity of any group of items parceled. Furthermore, this method is applied more often when a large group of items is being evaluated to assist in dimension reduction. In contradistinction, in our application with few items, we were careful to use this approach only where there was strong clinical evidence of a close relationship justifying parceling and/or allowing item residual covariances.

Historically, the expectation in CFA is for larger sample sizes, such as greater than 200,[62] which we meet. However, more recently, methodological research has shown that much smaller sample sizes may be adequate

for CFA.[63,64] Furthermore, in terms of invariance measurement, fixed standards are difficult to define due to variations in factor quality[65,66] that normally occur and the fact that invariance is an effect size of sorts that cannot be known ahead of time. Based on the relevant methodological literature, then, we expect that our sample sizes are more than adequate in both the case of CFA and invariance measurement.

A potential weakness of psychometric evaluations such as ours in indices such as the EQ-5D or the FACT-LUI is that preference-based measures are by necessity parsimonious, which goes against the conventional wisdom of psychometric evaluation where there are at least 3 or more items often used to separate error from variance and loadings. Thus, there is an element of uncertainty that is inherent to the comparison of formative versus reflective methodologies.

One might argue that having some patients in a formal trial in active therapy (group 2) and another group in which patients are in various points in their disease trajectory makes the 2 groups different and so not applicable to invariance measurement. We argue that they are much more similar than not and are in fact a good test of invariance. All are lung cancer patients and all come from large academic institutions; thus, they are very likely to have been a part of some type of trial at some point. Group 2 had an expected higher level of morbidity, which is apparent in the results, but any good measure of HrQoL should be robust to such circumstances. Finally, the fact that almost all have non–small-cell carcinoma and a small minority do not is irrelevant to concerns about group differences, since lung cancer HrQoL measures do not differentiate by tissue type.[33,67]

Conclusions

The FACT-LUI shows potential for diverse applications. A preference-based index showing such evidence of reasonable psychometrics at the CFA level suggests multiple applications in a small package. These capabilities further allow easy incorporation into clinical trials with a minimum of effect on the burden of a protocol. Such a measure is likely to be advantageous in current circumstances, where the trajectory of cancer care shows that economic analysis is critical for the future[68–70] alongside other patient-reported outcomes.

Acknowledgments

The authors acknowledge the valuable statistical assistance of Gregory R. Hancock, PhD. We also acknowledge the cited data set kindly provided by Susan E. Yount, PhD.

ORCID iD

J. Shannon Swan https://orcid.org/0000-0003-1338-9062

Data Availability

Data that support the conclusions of this article are included in a supplementary file in the form of the correlation matrices, means, and standard deviations used to generate the models. Also included are the unstandardized estimates from modeling and external correlate measures.

Supplemental Material

Supplementary material for this article is available online at https://doi.org/10.1177/23814683231186992.

References

1. Van Der Weijst L, Lievens Y, Schrauwen W, Surmont V. Health-related quality of life in advanced non-small cell lung cancer: a methodological appraisal based on a systematic literature review. *Front Oncol*. 2019;9:715. doi:10.3389/fonc.2019.00715
2. Kuon J, Blasi M, Unsold L, et al. Impact of molecular alterations on quality of life and prognostic understanding over time in patients with incurable lung cancer: a multicenter, longitudinal, prospective cohort study. *Support Care Cancer*. 2022;30(4):3131–40. doi:10.1007/s00520-021-06736-2
3. Huber RM, Kauffmann-Guerrero D, Hoffmann H, Flentje M. New developments in locally advanced non-small cell lung cancer. *Eur Respir Rev*. 2021;30(160):200227. doi:10.1183/16000617.0227-2020
4. Sullivan DR, Eden KB, Dieckmann NF, et al. Understanding patients' values and preferences regarding early stage lung cancer treatment decision making. *Lung Cancer*. 2019;131:47–57. doi:10.1016/j.lungcan.2019.03.009
5. Schnog JB, Samson MJ, Gans ROB, Duits AJ. An urgent call to raise the bar in oncology. *Br J Cancer*. 2021;125(11):1477–85. doi:10.1038/s41416-021-01495-7
6. National Institutes of Health. *Cancer Trends Progress Report*. Bethesda (MD): National Cancer Institute, NIH, HHS. October 2022. Available from: https://progressreport.cancer.gov
7. Fayers PM. Quality-of-life measurement in clinical trials—the impact of causal variables. *J Biopharm Stat*. 2004;14(1):155–76. doi:10.1081/bip-120028512
8. Richardson J. Psychometric validity and Multiattribute Utility (MAU) instruments. 2010 Research Paper 2010 (57), 1–19. Available from: https://www.semanticscholar.org/paper/Psychometric-Validity-and-Multi-Attribute-Utility-Richardson/c04bbc93c44eb5da4265b39661bbd9d58da72398
9. von Winterfeldt D, Edwards W. Structuring for decision analysis. In: von Winterfeldt D, Edwards W, eds. *Decision Analysis and Behavioral Research*. New York: Cambridge University Press; 1986. p 42–3.

10. Palta M, Chen HY, Kaplan RM, Feeny D, Cherepanov D, Fryback DG. Standard error of measurement of 5 health utility indexes across the range of health for use in estimating reliability and responsiveness. *Med Decis Making*. 2011;31(2):260–9. doi:10.1177/0272989X10380925
11. Cherepanov D, Palta M, Fryback DG. Underlying dimensions of the five health-related quality-of-life measures used in utility assessment: evidence from the National Health Measurement Study. *Med Care*. 2010;48(8):718–25. doi:10.1097/MLR.0b013e3181e35871
12. Dewitt B, Feeny D, Fischhoff B, et al. Estimation of a preference-based summary score for the patient-reported outcomes measurement information system: the PROMIS®-Preference (PROPr) scoring system. *Med Decis Making*. 2018;38(6):683–98. doi:10.1177/0272989X18776637
13. Feng YS, Jiang R, Kohlmann T, Pickard AS. Exploring the internal structure of the EQ-5D using non-preference-based methods. *Value Health*. 2019;22(5):527–36. doi:10.1016/j.jval.2019.02.006
14. 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):365–76. doi:10.1093/jnci/85.5.365
15. Cella DF, Tulsky DS, Gray G, et al. The functional assessment of cancer therapy scale: development and validation of the general measure. *J Clin Oncol*. 1993;11(3):570–9. doi:10.1200/JCO.1993.11.3.570
16. Ware JE Jr, Snow KK, Kosinski M, Gandek B. *SF-36 Health Survey: Manual and Interpretation Guide*. Boston (MA): The Health Institute, New England Medical Center; 1993.
17. Ware J Jr, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Med Care*. 1996;34(3):220–33. doi:10.1097/00005650-199603000-00003
18. DeVellis RF. *Scale Development: Theory and Applications*. 3rd ed. Thousand Oaks (CA): Sage; 2017.
19. Teckle P, Peacock S, McTaggart-Cowan H, et al. The ability of cancer-specific and generic preference-based instruments to discriminate across clinical and self-reported measures of cancer severities. *Health Qual Life Outcomes*. 2011;9:106. doi:10.1186/1477-7525-9-106
20. King MT, Norman R, Mercieca-Bebber R, et al. The Functional Assessment of Cancer Therapy Eight Dimension (FACT-8D), a Multi-Attribute Utility Instrument Derived From the Cancer-Specific FACT-General (FACT-G) Quality of Life Questionnaire: development and Australian Value Set. *Value Health*. 2021;24(6):862–73. doi:10.1016/j.jval.2021.01.007
21. King MT, Viney R, Simon Pickard A, et al. Australian Utility Weights for the EORTC QLU-C10D, a multi-attribute utility instrument derived from the Cancer-Specific Quality of Life Questionnaire, EORTC QLQ-C30. *Pharmacoeconomics*. 2018;36(2):225–38. doi:10.1007/s40273-017-0582-5
22. Kind P, Macran S. Eliciting social preference weights for functional assessment of cancer therapy-lung health states. *Pharmacoeconomics*. 2005;23(11):1143–53. doi:10.2165/00019053-200523110-00006
23. Swan JS, Lennes IT, Stump NN, et al. A patient-centered utility index for non-small cell lung cancer in the united states. *MDM Policy Pract*. 2018;3(2):2381468318801565. doi:10.1177/2381468318801565
24. Yount SE, Rothrock N, Bass M, et al. A randomized trial of weekly symptom telemonitoring in advanced lung cancer. *J Pain Symptom Manage*. 2014;47(6):973–89. doi:10.1016/j.jpainsymman.2013.07.013
25. Brown TA.Specification and interpretation of CFA models. In: Brown TA, ed. *Confirmatory Factor Analysis for Applied Research*. New York: Guilford; 2015. p 88–129.
26. Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Dev Rev*. 2016;41:71–90. doi:10.1016/j.dr.2016.06.004
27. Iyer S, Roughley A, Rider A, Taylor-Stokes G. The symptom burden of non-small cell lung cancer in the USA: a real-world cross-sectional study. *Support Care Cancer*. 2014;22(1):181–7. doi:10.1007/s00520-013-1959-4
28. Henoch I, Ploner A, Tishelman C. Increasing stringency in symptom cluster research: a methodological exploration of symptom clusters in patients with inoperable lung cancer. *Oncol Nurs Forum*. 2009;36(6):E282–92. doi:10.1188/09.ONF.E283-E292
29. Chen E, Nguyen J, Cramarossa G, et al. Symptom clusters in patients with lung cancer: a literature review. *Expert Rev Pharmacoecon Outcomes Res*. 2011;11(4):433–9. doi:10.1586/erp.11.56
30. Yang P, Cheville AL, Wampfler JA, et al. Quality of life and symptom burden among long-term lung cancer survivors. *J Thorac Oncol*. 2012;7(1):64–70. doi:10.1097/JTO.0b013e3182397b3e
31. Pirri C, Bayliss E, Trotter J, et al. Nausea still the poor relation in antiemetic therapy? The impact on cancer patients' quality of life and psychological adjustment of nausea, vomiting and appetite loss, individually and concurrently as part of a symptom cluster. *Support Care Cancer*. 2013;21(3):735–48. doi:10.1007/s00520-012-1574-9
32. Brazier J, Czoski-Murray C, Roberts J, Brown M, Symonds T, Kelleher C. Estimation of a preference-based index from a condition-specific measure: the King's Health Questionnaire. *Med Decis Making*. 2008;28(1):113–26. doi:10.1177/0272989X07301820
33. Cella DF, Bonomi AE, Lloyd SR, Tulsky DS, Kaplan E, Bonomi P. Reliability and validity of the Functional Assessment of Cancer Therapy-Lung (FACT-L) quality of life instrument. *Lung Cancer*. 1995;12(3):199–220. doi:10.1016/0169-5002(95)00450-f
34. Eton DT, Cella D, Yount SE, Davis KM. Validation of the functional assessment of cancer therapy—lung symptom index-12 (FLSI-12). *Lung Cancer*. 2007;57(3):339–47. doi:10.1016/j.lungcan.2007.03.021

35. Cheville AL, Novotny PJ, Sloan JA, et al. Fatigue, dyspnea, and cough comprise a persistent symptom cluster up to five years after diagnosis with lung cancer. *J Pain Symptom Manage*. 2011;42(2):202–12. doi:10.1016/j.jpainsymman.2010.10.257
36. Carnio S, Di Stefano RF, Novello S. Fatigue in lung cancer patients: symptom burden and management of challenges. *Lung Cancer (Auckl)*. 2016;7:73–82. doi:10.2147/LCTT.S85334
37. Little TD, Rhemtulla M, Gibson K, Schoemann AM. Why the items versus parcels controversy needn't be one. *Psychol Methods*. 2013;18(3):285–300. doi:10.1037/a0033266
38. Card NA.Coding individual studies. In: Card NA, ed. *Applied Meta-Analysis for Social Science Research*. New York: Guilford Publications; 2012. p 148.
39. Little TD.Overview and foundations of SEM. In: Little TD, ed. *Longitudinal Structural Equation Modeling*. New York: Guilford Publications; 2013. p 1–36.
40. Krouwer JS, Monti KL. A simple, graphical method to evaluate laboratory assays. *Eur J Clin Biochem*. 1995;33(8):525–7.
41. Barsevick A. Symptom clusters. *Semin Oncol Nurs*. 2016;32(4):333. doi:10.1016/j.soncn.2016.09.001
42. Grillo L. A possible role of anhedonia as common substrate for depression and anxiety. *Depress Res Treat*. 2016;2016:1598130. doi:10.1155/2016/1598130
43. Moller HJ, Bandelow B, Volz HP, Barnikol UB, Seifritz E, Kasper S. The relevance of 'mixed anxiety and depression' as a diagnostic category in clinical practice. *Eur Arch Psychiatry Clin Neurosci*. 2016;266(8):725–36. doi:10.1007/s00406-016-0684-7
44. Pitman A, Suleman S, Hyde N, Hodgkiss A. Depression and anxiety in patients with cancer. *BMJ*. 2018;361:k1415. doi:10.1136/bmj.k1415
45. Mineka S, Watson D, Clark LA. Comorbidity of anxiety and unipolar mood disorders. *Annu Rev Psychol*. 1998;49:377–412. doi:10.1146/annurev.psych.49.1.377
46. Kasper S. Depression and anxiety-separate or continuum? *World J Biol Psychiatry*. 2001;2(4):162–3. doi:10.3109/15622970109026804
47. Kessler RC, DuPont RL, Berglund P, Wittchen HU. Impairment in pure and comorbid generalized anxiety disorder and major depression at 12 months in two national surveys. *Am J Psychiatry*. 1999;156(12):1915–23. doi:10.1176/ajp.156.12.1915
48. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62(6):617–27. doi:10.1001/archpsyc.62.6.617
49. Kroenke K, Spitzer RL, Williams JB, Lowe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics*. 2009;50(6):613–21. doi:10.1176/appi.psy.50.6.613
50. Gift AG, Jablonski A, Stommel M, Given CW. Symptom clusters in elderly patients with lung cancer. *Oncol Nurs Forum*. 2004;31(2):202–12. doi:10.1188/04.ONF.202-212
51. Raykov T.Fundamentals of structural equation modeling. In: Raykov T, Marcoulides GA, eds. *A First Course in Structural Equation Modeling*. Mahwah (NJ): Lawrence Erlbaum Associates, Inc.; 2006. p 31.
52. Raykov T.Scale construction and development using structural equation modeling. In: Hoyle RH, ed. *Handbook of Structural Equation Modeling*. New York: Guilford Publications; 2012. p 472–92.
53. Rigdon EE.Structural equation modeling. In: Marcoulides GA, ed. *Modern Methods for Business Research*. Mahwah (NJ): Lawrence Erlbaum; 1998. p 251–94.
54. Rhemtulla M, Brousseau-Liard P, Savalei V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol Methods*. 2012;17(3):354–73. doi:10.1037/a0029315
55. Satorra A, Bentler PM. Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*. 2010;75(2):243–8. doi:10.1007/s11336-009-9135-y
56. Kelloway EK.Assessing model fit. In: Kelloway EK, ed. *Using MPlus for Structural Equation Modeling*. Thousand Oaks (CA): Sage; 2015. p 21–36.
57. Kline RB.Estimation and local fit testing. In: Kline RB, ed. *Principles and Practice of Structural Equation Modeling*. New York: Guilford Publications; 2016. p 231–60.
58. Hu L, Bentler PM. Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model*. 1999;6:1–55. doi:10.1080/10705519909540118
59. American Cancer Society. *Cancer Facts and Figures*. Atlanta (GA): American Cancer Society; 2020.
60. Hanmer J, Cella D, Feeny D, et al. Selection of key health domains from PROMIS® for a generic preference-based scoring system. *Qual Life Res*. 2017;26(12):3377–85. doi:10.1007/s11136-017-1686-2
61. Matsunaga M. Item parceling in structural equation modeling: a primer. *Commun Methods Meas*. 2008;2(4):260–93. doi:10.1080/19312450802458935
62. Barrett P. Structural equation modeling: adjudging model fit. *Pers Individ Differ*. 2007;42:815–24.
63. Marsh HW, Hau KT, Balla JR, Grayson D. Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behav Res*. 1998;33(2):181–220. doi:10.1207/s15327906mbr3302_1
64. Gagne P, Hancock GR. Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behav Res*. 2006;41(1):65–83. doi:10.1207/s15327906mbr4101_5
65. Hancock GR, Mueller RO. The reliability paradox in assessing structural relations within covariance structure models. *Educ Psychol Meas*. 2011;71(2):306–24. doi:10.1177/0013164410384856
66. McNeish D, An J, Hancock GR. The thorny relation between measurement quality and fit index cutoffs in latent variable models. *J Pers Assess*. 2018;100(1):43–52. doi:10.1080/00223891.2017.1281286

67. Bergman B, Aaronson NK, Ahmedzai S, Kaasa S, Sullivan M. The EORTC QLQ-LC13: a modular supplement to the EORTC Core Quality of Life Questionnaire (QLQ-C30) for use in lung cancer clinical trials. EORTC Study Group on Quality of Life. *Eur J Cancer*. 1994;30A(5):635–42. doi: 10.1016/0959-8049(94)90535-5
68. Shankaran V. Cost considerations in the evaluation and treatment of colorectal cancer. *Curr Treat Options Oncol*. 2015;16(8):41. doi:10.1007/s11864-015-0354-4
69. Kelly R, Houseknecht S. Considerations in the care of non-small-cell lung cancer: the value imperative. *Oncology (Williston Park)*. 2018;32(11):534–40.
70. Greenberg D, Earle C, Fang CH, Eldar-Lissai A, Neumann PJ. When is cancer care cost-effective? A systematic overview of cost-utility analyses in oncology. *J Natl Cancer Inst*. 2010;102(2):82–8. doi:10.1093/jnci/djp472