Reducing patient burden to the FACT-Melanoma quality of life questionnaire

Richard J. Swartz, Ph.D.[†], **George P. Baum, M.S.**[+], **Robert L. Askew, M.P.H.**[‡], **J. Lynn Palmer, Ph.D.**[¥], **Merrick I. Ross, M.D.**[‡], and **Janice N. Cormier, M.D., M.P.H.**[‡]

[†]Jones Graduate School of Business, Rice University, Houston, Texas; 77005 phone: 713-348-6309, fax 713-328-6296, rswartz@rice.edu [+]Department of Behavioral Science, Unit 1330, The University of Texas M. D. Anderson Cancer Center, 1155 Hermann Pressler Drive, Houston, T 77030-3721, 713-745-1199 gpbaum@mdanderson.org [‡]Department of Surgical Oncology, Unit 444, The University of Texas MD Anderson Cancer Center, 1500 Holcombe Boulevard, Houston, TX 77030-4009; Phone: 713-792-6949, Fax: 713-745-1921. jcormier@mdanderson.org, mross@mdanderson.org [¥]Department of Biostatistics, Unit 1411, The University of Texas M. D. Anderson Cancer Center, 1400 Hermann Pressler Drive, Houston, Texas 77030-4008, 713-792-7570l palmer@mdanderson.org

Abstract

Objectives—Respondent burden has been defined as the cumulative demand placed on study participants related to the use of questionnaires or measurement instruments. The aim of this study was to reduce respondent burden associated with the FACT-Melanoma (FACT-M), a melanoma-specific quality of life questionnaire, through item reduction using multiple psychometric approaches.

Methods—Data for this study were pooled from three IRB-approved protocols. Poorly performing items were identified through distributional and correlation analyses, confirmatory factor analysis, reliability estimation, and Rasch-based approaches in a developmental dataset, and the reduced scale was assessed in a separate testing cohort. Validity, reliability, goodness of fit, and Rasch-based testing were carried out for both the full and reduced scales.

Results—The clinical characteristics of the development (n=198) and testing (n=204) cohorts were similar. Three items identified through classical psychometric approaches and 3 items identified by Rasch-based analyses were dropped from the FACT-M subscale. Two additional items were identified for potential reduction but were ultimately maintained due to the adverse consequences to the psychometric integrity of the reduced instrument.

Conclusions—The reduced FACT-M module contains 18 items. In addition to psychometric assessment, expert consultation was essential when examining areas of content redundancy and was critical when considering specific items for removal. This methodological approach reduced respondent burden by 25% while maintaining the psychometric integrity of the FACT-M.

Keywords

FACT-Melanoma questionnaire; respondent burden; item reduction analysis

Introduction

Health-related quality of life (HRQOL) questionnaires are multidimensional instruments used to assess domains related health status [1]. The Functional Assessment of Chronic Illness and Therapy is an outcome measurement system composed of a core questionnaire for cancer paitents, the Functional Assessment of Cancer Therapy-General (FACT-G), and a disease-specific module [2-4]. The FACT-Melanoma (FACT-M) includes a melanoma module comprised of 24 total items (www.facit.org) which can be administered as 2 separate subscales (i.e. the Melanoma Subscale [MS] and the Melanoma Surgical Subscale [MSS]). The MS contains 16 items assessing general HRQOL in 3 domains (physical well-being: 12 items, emotional well-being: 3 items, social well-being: 1 item). The MSS assesses surgery-specific concerns with 8 items related to treatment of the primary tumor site, which is particularly relevant to patients with localized disease (stage I/II) [5]. The summary score from the FACT-M module has been validated as an independent measure or can be combined with the FACT-G as a measure of overall cancer-related QOL [6].

While the FACT-M has demonstrated strong internal consistency and high test-retest reliability, [6] issues related to patient respondent burden have been increasingly acknowledged [7]. Respondent burden has been defined as the cumulative demand placed on study participants related to the use of questionnaires or surveys [7]. Respondent burden can be particularly problematic for patients who are incapacitated due to advanced disease or who are subjected to repeated survey administration. Less complex quality of life questionnaires that require minimal time for completion are highly desirable[8, 9] as excessive imposed burden on clinical trial participants can undermine both response rates and data quality [2, 8, 10, 11]. In order to increase the clinical utility of the FACT-M, our group sought to lower respondent burden by reducing the number of items while maintaining the psychometric integrity of the instrument.

Methods

Quality of life data collected using the FACT-M were pooled from three institutional review board-approved prospective studies. The first study (2006-present) involves the collection of FACT-M data with assessments of limb volume over time for melanoma. The second group includes melanoma patients enrolled in a study (2008-2009) to assess immune responses to peptide vaccines. The third dataset includes melanoma patients (2004-2006) who participated in the validation study of the FACT-M. The data from these three studies were pooled and then distributed randomly into a developmental and validation dataset with equal distribution of stages I/II, stage III, and stage IV patients. The validation cohort was used to compare and confirm the performance of the full and reduced models.

Because the FACT-G and the melanoma module assess overlapping domains, the first phase of analyses aimed to identify and eliminate redundancies and poorly performing items. Pearson's bivariate correlation analysis was used to identify items from the melanoma module that were highly correlated (and therefore redundant) with the FACT-G items. Items

with statistically significant correlations of 0.7 or greater were dropped. In addition, univariate analysis identified items from the MS and MSS that exhibited extremely low variability as an indicator of a poorly discriminating item.

The second phase of the analysis explored the possibility of shortening the instrument using both classical and more modern psychometric techniques. First, a confirmatory factor analysis using the development dataset was performed to confirm the domain structure identified in the previous FACT-M validation study [5]. A comparative fit index (CFI) value of ≥ .95 [12] and a root mean square error of approximation (RMSEA) fit statistic value of ≤ .05 [13] were considered indicative of good fit. Next, a Rasch measurement based approach was performed using WINSTEPS (version 3.66) software.[14] Dimensionality of the underlying traits was assessed, and if unidimensionality for the full FACT-M scale did not hold sufficiently for Rasch Analysis, the melanoma (MS) and melanoma surgical subdomains (MSS) would be analyzed separately. We fit a partial credit model, which is a widely used Rasch model for polytomous items such as those encountered on the FACT-M, to allow the thresholds between categories to be item specific -- that is, the relative distance on the latent scale for one item was not required to be uniform across all items [15]. Infit and outfit mean squares were used to assess item fit, as the infit statistic is more sensitive to unexpected behavior of items located near the subjects HRQOL, while the outfit statistic is sensitive to unexpected behavior of items that are located far from the subject's quality of life level. Typically, items with infit or outfit statistics between 0.7 and 1.3 are maintained, [16] however since this scale ultimately was not scored according to Rasch scoring and there were competing considerations such as item reliability or content coverage, our assumption was that an item could be retained even if the infit or outfit statistic was as high as 2.0. Reliability and separation indices were also used to evaluate the Rasch measure. If dropping the item resulted in more than 20% of the sample having extreme scores indicating a ceiling or floor effect, the item was retained regardless of fit.

After removing poorly performing items, the reduced scale was tested using the validation dataset. Goodness of fit was again assessed using the CFI and RMSEA. Internal reliability was evaluated using Cronbach's α for both the full and reduced scales. Rasch analysis was performed on the validation dataset for the full and reduced scales to assess robustness of fit.

Results

From the combined groups of patients (n=458), 56 were excluded from the item-reduction analysis due to incomplete survey responses. The analytic cohort (n=402) was predominately Caucasian (97%) with a slight majority of males (56%). Similar distributions across demographic and clinical characteristics were observed for respondents in the developmental (n=198) and testing (n=204) datasets (Table 1).

The initial phase of analysis revealed that item 16 (I feel fatigued) of the melanoma module was highly correlated ($r = .703$ p < .001) with item 1 of the FACT-G (I have a lack of energy). Item 16 was also noted to have a high percentage of missing values (37%), and was dropped prior to analysis due to redundancy. Items 7 (I have had fevers) and 11 (I have noticed blood in my stool) of the melanoma module exhibited extremely low variability with 96% and 99% of subjects responding "not at all" for the two items, respectively. Due to this low variability, items 7 and 11 were also dropped prior to the next phase of analysis.

Confirmatory factor analysis showed that a three factor model (one factor for the MSS and two correlated physical and psychosocial factors comprising the MS) resulted in poor fit statistics (chi Square: 153.2, df = 76, p< .001; CFI = .791; RMSEA = .077). When performing a Rasch analysis, principal component analysis of the residuals revealed that

there was a strong secondary factor present in the data (3.6 eigen value units, which is greater than the 1.6 eigen value units expected if the data were unidimensional). This finding, along with support from the scoring manuals of the FACT-M recommending separate subscale scoring, led to separate analyses of the MS and MSS to satisfy the unidimensionality assumption necessary for Rasch analysis.

Rasch analysis of the MS revealed that item 9 (I have a good appetite) had disordered response categories, which violates the intended property of the item and the scale, so it was dropped. Item 3 (I worry about the appearance of surgical scars) also exhibited disordered categories, but collapsing the lowest two response categories into one ameliorated this finding, so the item was maintained. Item 10 (I have aches and pains in my bones) had high infit and outfit statistics (1.47 and 1.55, respectively) and was also dropped. Dropping any additional items resulted in more than 20% of subjects in the training dataset having extreme scores, and the infit and outfit statistics of the remaining items were all less than 2.0 (Table 2).

Rasch analysis of the MSS revealed that item 24 (I have a good range of motion in my arm or leg), a reverse coded item, had disordered categories. As disordered categories violate scale and item assumptions, this item was dropped. Item 23 (I feel numbness at my surgical site) was identified through Rasch analysis fit statistics. Principal component analysis of the residuals indicated that this item fit a separate dimension from the other items. However, dropping this item resulted in over 50% of respondents having extreme scores for this subscale. This item also behaved well in terms of the category response frequencies and the Rasch category thresholds, so item 23 was maintained. In total, 6 items (items 7, 9, 10, 11, 16, and 24) were dropped from the scales. Confirmatory factor analysis was used to verify that the factor structure of the original melanoma module was maintained in the reduced scale with acceptable measures of fit from the training data (chi-square = 268.9, p < .001, CFI =.975 and RMSEA = 0.078).

The final phase of the study involved assessing the performance of the full and reduced scales on the validation dataset. Confirmatory factor analysis revealed that the melanoma specific items of the reduced FACT-M exhibited acceptable fit (chi-square value of 224.0, p < .001, CFI = .981 and RMSEA = .070) and better fit than the original melanoma specific items. Reliability estimates were also compared between the full and reduced scales (Table 3). Cronbach's α values for the MS and the MSS were excellent for the full and reduced scales (all ≥ 0.90). Reliability remained high for the FACT-M total scores, defined as the FACT-G items scored with the MS and MSS subscales of the FACT-M module ($\alpha = 0.97$ for the full and reduced scales). The item reliability dropped from 0.96 to 0.79 in the MS and from 0.92 to 0.78 in the MSS. While the fit statistics are noticeably higher for items 2 and 3 and are within acceptable limits for item 10 (which was dropped for the training data), overall Rasch fit statistics from the validation dataset were within acceptable limits and confirm the results from the training data. Correlations between the full and reduced FACT-M were very high (r = .997 and .996 for the training and validation datasets, respectively), confirming criterion validity of the reduced instrument. Figure 1 presents multiple scatter plots comparing the performance of the full and reduced scales in multiple contexts (i.e. combined melanoma module, MS, MSS, and total FACT-M). These plots demonstrate that the score distributions for the full and reduced scales are very similar.

Discussion

The reduced FACT-M retains 18 of the original 24-items without compromising the psychometric integrity of the original instrument; the reliability and factor structure of the original FACT-M were maintained as was the content validity. The method employed was

an iterative process that entailed examining both the performance characteristics of individual items and the effects of removing specific items on the overall functioning of the instrument. Our methods capitalized on the strengths of multiple psychometric approaches. For example, Rasch models assume that a single, unidimensional trait drives item response. We were able to assess this assumption before implementing Rasch analysis using classical test theory (CTT). A typical test for unidimensionality in Rasch analysis uses CTT by performing a principal component analysis on the residuals (after the primary dimension is removed). If the residual contrasts identify non-random factors, this is evidence of multidimensionality. Another important strength of the analysis presented lies in the validation process which was carried out in a separate cohort of patients – an important but often overlooked step that produces better measures of construct validity and reliability [17].

In addition to the assessments of data dimensionality, the statistical methods of CTT offer several key strengths. Compared to item-response models like those in Rasch analysis, CTT models are less mathematically complex, and as such, it is easier to satisfy the statistical assumptions required for analysis and to overcome traditional weaknesses associated with moderate sample sizes [18]. CTT also facilitates straightforward analysis of precision with separate summary estimates for the observed scores and the expected level of error from their "true" scores that are easily interpretable [18].

Although some consider Rasch measurement a type of Item Response Theory (IRT) approach, there are those that argue it is a separate measurement philosophy with its own benefits [19-21]. In either case, a Rasch measurement approach offers several advantages over the classical psychometric strategies. Specifically, Rasch measurement models assess item and person scores independently, calibrate them to the same scale, and provide model-driven fit statistics for individual items that can be used to identify poorly performing items [19, 21, 22]. The reliability index is analogous to Cronbach's alpha and measures the reproducibility of the person's scores. Although the item reliability index has no analog in classical psychometrics, it has an interpretation similar to the person reliability; item reliability is a measure of the reproducibility of the item parameter estimates. Items are generally considered poorly fitting when they have infit or outfit statistics larger than 1.3 or lower than 0.7 [16].

The Rasch analysis revealed that the item reliability, although relatively high for the full scale, reduces to more moderate values for the shortened scale. Further inspection, however, revealed that the high item reliability for the full scale is strongly influenced by a poorly performing item in both subscales (items 9 and 24). Because these items were both reverse-coded and scores did not increase monotonically as categories increased, these two items may have artificially inflated the item reliability of the original scale.

Through this item reduction process, each of the HRQOL domains was maintained in the reduced instrument. Others have similarly found that an iterative and multifaceted approach can lead to substantial reduction in respondent burden without major sacrifices in the psychometric integrity of the instrument.[23-26] Mead et al employed elements of CTT and Rasch analysis when identifying items for potential reduction from a psychological measure of hypomania [25]. More directly related to cancer, a recent CTT-driven analysis was successfully undertaken for the 11-item FACT/Gynecologic Oncology Group Neurotoxicity subscale (FACT/GOG-Ntx) to reduce respondent burden for endometrial cancer patients, resulting in a substantially reduced 4-item short form [24]. Likewise, Prieto and colleagues in a study of HRQOL compared the results of separate CTT and Rasch-based analyses reducing a generic measure of health status to a short-form version [26].

Conclusion

Reducing respondent burden by shortening established quality of life instruments is a multi-phased methodological process. Comprehensive assessment of patient HRQOL is particularly important in the context of clinical research, where the burden of participation may vary by intensity and degree according to the subject's condition, prognosis, mental state, and support systems [27]. Shorter and less complex HRQOL instruments that maintain sufficient psychometric integrity for research purposes are therefore more desirable, particularly given that excessive burden can undermine questionnaire response and data quality [8].

Acknowledgments

Abbreviations

CFI	Comparative fit index
CTT	Classical test theory
FACT-G	Functional Assessment of Cancer Therapy - General
FACT-M	Functional Assessment of Cancer Therapy – Melanoma
HRQOL	Health-related quality of life
MS	Melanoma Subscale
MSS	Melanoma Surgical Subscale
RMSEA	Root mean square error of approximation

References

1. Langenhoff BS, Krabbe PF, Wobbes T, Ruers TJ. Quality of life as an outcome measure in surgical oncology. Br J Surg. 2001; 5:643–52. [PubMed: 11350434]

2. Turner RR, Quittner AL, Parasuraman BM, Kallich JD, Cleeland CS, Mayo FDA Patient Reported Outcomes Consensus Meeting Group. et al. Patient-reported outcomes: instrument development and selection issues. Value in Health. 2007:S86–93. [PubMed: 17995478]

3. Webster K, Cella D, Yost K. The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System: properties, applications, and interpretation. Health Qual Life Outcomes. 2003:79. [PubMed: 14678568]

4. Cella D, Nowinski CJ. Measuring quality of life in chronic illness: the functional assessment of chronic illness therapy measurement system. Arch Phys Med Rehabil. 2002; 12(Suppl 2):S10–7. [PubMed: 12474167]

5. Cormier JN, Davidson L, Xing Y, Webster K, Cella D. Measuring quality of life in patients with melanoma: development of the FACT-melanoma subscale. The journal of supportive oncology. 2005; 2:139–45. [PubMed: 15796446]

6. Cormier JN, Ross MI, Gershenwald JE, Lee JE, Masnsfield PF, Camacho LH, et al. Prospective assessment of the reliablity, validity, and responsiveness to change of the Functional Assessment of

CancerTherapy-Melanoma (FACT-Melanoma) questionnaire. Cancer. 2008; 10:2249–2257. [PubMed: 18383513]

7. Science Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: Attributes and review criteria. Quality of Life Research. 2002; 3:193–205.

8. Bernhard J, Cella DF, Coates AS, Fallowfield L, Ganz PA, Moinpour CM, et al. Missing quality of life data in cancer clinical trials: serious problems and challenges. Stat Med. 1998; 5-7:517–32. [PubMed: 9549801]

9. Ballatori E. Unsolved problems in evaluating the quality of life of cancer patients. Ann Oncol. 2001:S11–3. [PubMed: 11804377]

10. Hays RD, Merz JF, Nicholas R. Response burden, reliability, and validity of the CAGE, Short MAST, and AUDIT alcohol screening measures. Behavior Research Methods, Instruments, & Computers. 1995; 2:277–280.

11. Turk DC, Dworkin RH, Burke LB, Gershon R, Rothman M, Scott J, et al. Developing patient-reported outcome measures for pain clinical trials: IMMPACT recommendations. Pain. 2006:208–215. [PubMed: 17069973]

12. Hu, L-t; Bentler, PM. Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. Structural Equation Modelling. 1999; 1:1–55.

13. MacCallum RC, Kim C, Marlarkey WB, Kiecolt-Glaser JK. Studying multivariate change using multilevel models and latent curve models. Multivariate Behavioral Research. 1997:215–253.

14. Linacre JM. Winsteps (R) (Version 3.66.0) [Computer Software]. 2009; 15 Masters GN. A Rasch Model for Partial Credit Scoring. Psychometrika. 1982; 2:149–174.

16. Wright, BD.; Stone, MH. Best test design. 1979.

17. Coste J, Guillemin F, Pouchot J, Fermanian J. Methodological approaches to shortening composite measurement scales. J Clin Epidemiol. 1997; 3:247–52. [PubMed: 9120523]

18. De Champlain AF. A primer on classical test theory and item response theory for assessments in medical education. Med Educ. 2010; 1:109–17. [PubMed: 20078762]

19. Andrich, D. Rasch models for measurement. 1988.

20. Baker JG. A comparison of graded response and rasch partial credit models with subjective well-being. J Educational and Behavorial Statistics. 2000; 3:253–270.

21. Bond, TG. Applying the Rasch model: fundamental measurement in the human sciences. 2001.

22. Cole JC, Rabin AS, Smith TL, Kaufman AS. Development and validation of a Rasch-derived CES-D short form. Psychol Assess. 2004; 4:360–72. [PubMed: 15584795]

23. Hays RD, Brown J, Brown LU, Spritzer KL, Crall JJ. Classical test theory and item response theory analyses of multi-item scales assessing parents' perceptions of their children's dental care. Med Care. 2006; 11(Suppl 3):S60–8. [PubMed: 17060837]

24. Huang HQ, Brady MF, Cella D, Fleming G. Validation and reduction of FACT/GOG-Ntx subscale for platinum/paclitaxel-induced neurologic symptoms: a gynecologic oncology group study. Int J Gynecol Cancer. 2007; 2:387–93. [PubMed: 17362317]

25. Meads DM, Bentall RP. Rasch analysis and item reduction of the hypomanic personality scale. Personality and Individual Differences. 2008; 8:1772–1783.

26. Prieto L, Alonso J, Lamarca R. Classical Test Theory versus Rasch analysis for quality of life questionnaire reduction. Health Qual Life Outcomes. 2003:27. [PubMed: 12952544]

27. Ulrich CM, Wallen GR, Feister A, Grady C. Respondent burden in clinical research: when are we asking too much of subjects? IRB. 2005; 4:17–20. [PubMed: 16220630]
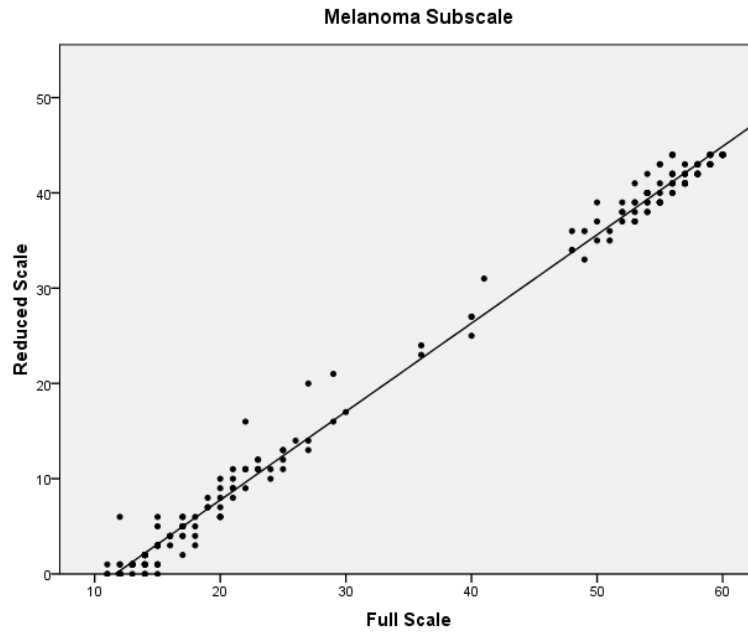
Figure 1.
Scatter plots of full vs. reduced FACT-Melanoma scores with linear predictions for the (a)
melanoma subscale, (b) melanoma surgical subscale, (c) combined melanoma module, and
(d) total FACT-Melanoma.

Figure 2.

Combined Melanoma Module
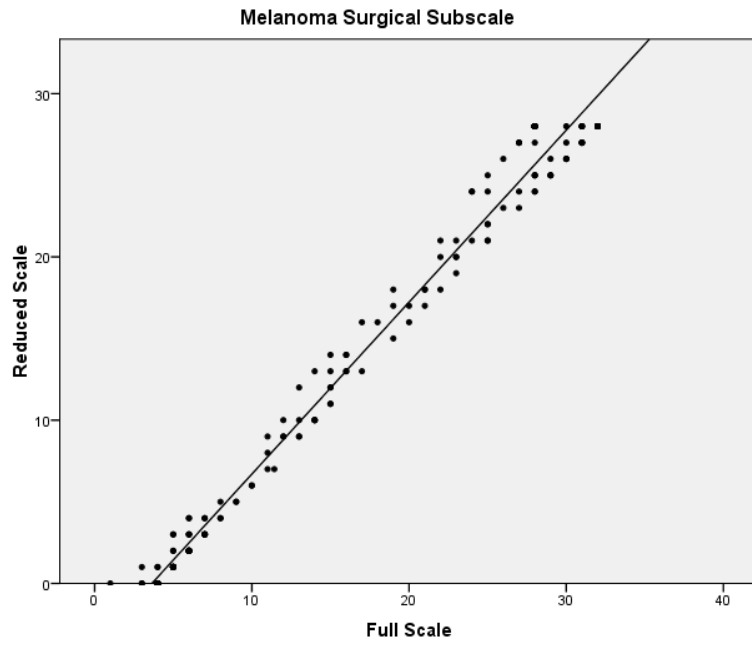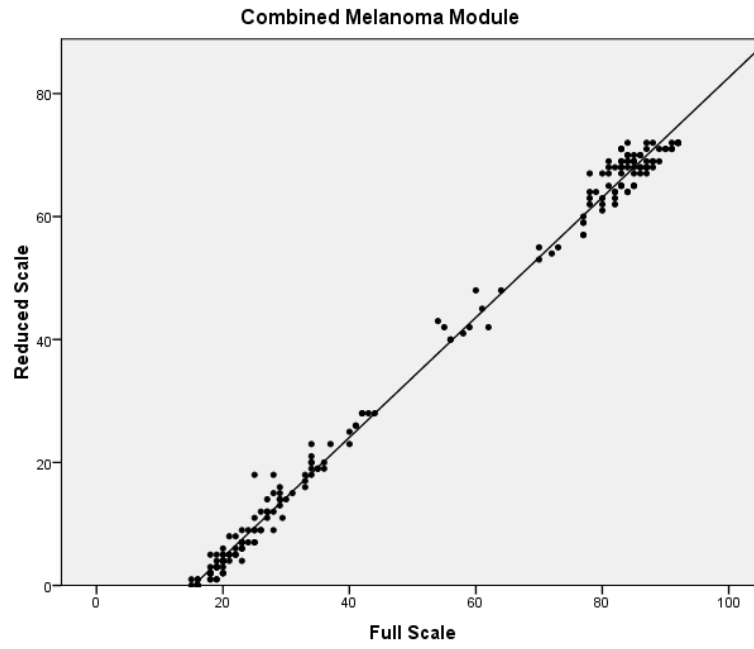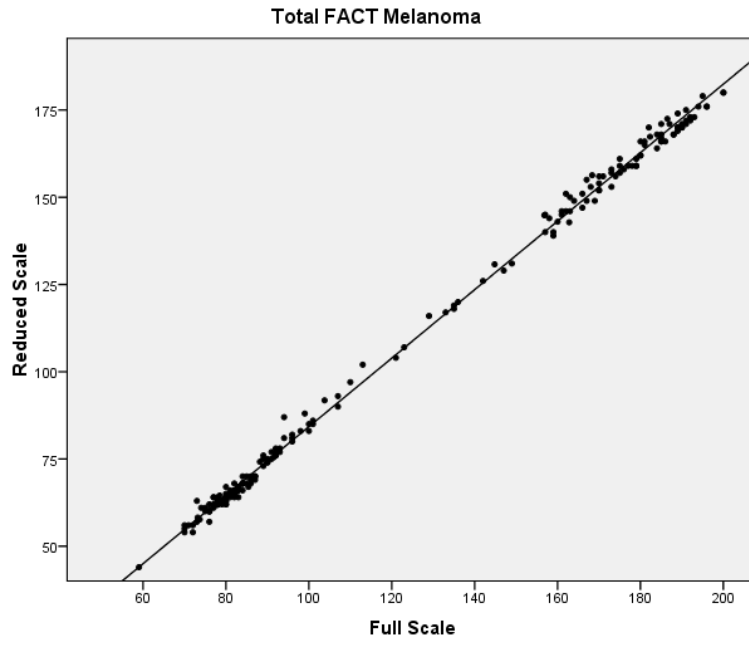
Figure 3.

Figure 4.

Table 1

Demographic and clinical characteristics of patient cohort stratified by dataset

		Training Dataset (n = 198)		Validation Dataset (n = 204)	
		n	%	n	%
Race	African-American	1	0.5	1	0.5
	Hispanic	1	0.5	5	2.5
	White, non-Hispanic	190	96.0	196	96.1
	Other	-	-	1	0.5
	Missing	6	3.0	1	0.5
Sex	Female	80	40.4	90	44.1
	Male	112	56.6	113	55.4
	Missing	6	3.0	1	0.5
Disease Stage	Stage I	33	16.7	53	26.0
	Stage II	33	16.7	41	20.1
	Stage III	66	33.3	92	45.1
	Stage IV	66	33.3	18	8.8

Table 2

FACT-M item scale scores, fit statistics, and model selection

Item	Domain	Final model	Training dataset Rasch scale Score	SE	Infit MNSQ	ZSTD	Outfit MNSQ	ZSTD	r	Validation dataset Rasch scale Score	SE	Infit MNSQ	ZSTD	Outfit MNSQ	ZSTD	r
Melanoma Subscale																
1	Physical	Maintained	−0.01	0.10	1.58	3.70	1.58	2.80	0.82	−0.06	0.10	0.81	−1.60	0.78	−1.50	0.90
2	Physical	Maintained	−0.19	0.10	1.31	1.70	1.47	1.60	0.84	−0.31	0.10	1.50	2.90	1.87	3.10	0.85
3	Physical	Maintained	0.32	0.12	1.12	0.80	1.61	2.10	0.82	0.67	0.12	1.55	3.30	1.99	3.60	0.81
4	Physical	Maintained	0.01	0.10	1.02	0.20	0.92	−0.20	0.86	0.05	0.12	0.78	−1.00	0.94	−0.10	0.88
5	Physical	Maintained	−0.36	0.09	0.92	−0.60	0.84	−1.10	0.87	−0.28	0.10	1.01	0.10	0.94	−0.30	0.88
6	Physical	Maintained	0.15	0.10	1.20	1.20	1.16	0.70	0.85	0.09	0.11	1.06	0.50	1.09	0.50	0.88
7	Physical	Dropped	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	Physical	Maintained	−0.12	0.10	0.70	−1.50	0.71	−0.60	0.86	−0.06	0.12	0.55	−2.30	0.56	−1.30	0.88
9	Physical	Dropped	−1.62	0.09	2.41	7.60	9.90	9.90	−0.04	−1.65	0.09	2.58	8.00	9.90	9.90	−0.10
10	Physical	Dropped	−0.24	0.09	1.47	3.30	1.55	2.90	0.80	0.06	0.08	0.82	−1.50	0.90	−0.60	0.84
11	Physical	Dropped	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	Social	Maintained	−0.10	0.10	0.96	−0.20	0.81	−0.90	0.87	−0.16	0.10	0.85	−0.90	0.69	−1.20	0.89
13	Emotional	Maintained	0.03	0.10	0.83	−1.30	0.84	−1.00	0.88	0.05	0.10	0.90	−0.90	0.95	−0.30	0.89
14	Emotional	Maintained	0.33	0.11	0.32	−4.30	0.40	−2.10	0.88	0.12	0.12	0.83	−0.80	0.61	−1.00	0.88
15	Emotional	Maintained	−0.06	0.10	0.81	−1.40	0.82	−1.10	0.88	−0.11	0.10	0.92	−0.60	0.98	−0.10	0.89
16	Physical	Dropped	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Melanoma Surgical Subscale																
17	Physical	Maintained	0.09	0.11	0.70	−1.90	0.56	−1.40	0.91	0.22	0.10	0.89	−0.70	0.94	−0.10	0.88
18	Physical	Maintained	−0.06	0.12	0.77	−1.60	0.76	−1.40	0.93	−0.14	0.10	0.80	−1.50	0.78	−1.20	0.90
19	Physical	Maintained	0.04	0.11	0.47	−4.00	0.35	−3.40	0.93	0.00	0.10	0.51	−4.10	0.37	−3.50	0.91
20	Physical	Maintained	0.23	0.12	0.61	−2.40	0.47	−2.00	0.91	0.18	0.10	0.43	−4.90	0.30	−3.80	0.91
21	Physical	Maintained	0.22	0.12	0.41	−3.80	0.23	−3.00	0.92	0.17	0.10	0.41	−4.90	0.35	−2.80	0.90
22	Physical	Maintained	−0.09	0.11	0.75	−1.50	0.40	−1.70	0.90	0.11	0.10	0.55	−3.50	0.37	−2.70	0.90
23	Physical	Maintained	−0.43	0.12	2.72	9.00	2.66	8.70	0.84	−0.54	0.11	2.81	9.90	2.80	9.90	0.77
24	Physical	Dropped	−0.81	0.06	2.96	9.90	9.90	9.90	−0.05	−0.67	0.06	3.25	9.90	9.90	9.90	−0.16

Abbreviations: SE, standard error; MNSQ, mean square value of the fit statistic; ZSTD, standardized fit statistic; r, Pearson's correlation coefficient between item score and the total score

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Reliability[a] measures for the full and reduced FACT-M subscales

	Total FACT-M		Melanoma Module		Melanoma Subscale		Melanoma Surgical Subscale	
	Full	Reduced	Full	Reduced	Full	Reduced	Full	Reduced
No. of items	51	45	24	18	16	11	8	7
Cronbach's alpha	0.97	0.97	0.97	0.99	0.96	0.98	0.90	0.97
Person reliability	--	--	--	--	0.88	0.88	0.72	0.82
Item reliability	--	--	--	--	0.96	0.79	0.92	0.78

[a]Item 16 not included in the reliability analysis due to missing values, and person and item reliability are calculated on the full scale with 13 items after dropping items 7, 11, and 16 in the initial phase of the analysis. They are not calculated for the Total FACT-M nor for the Melanoma Module, because these estimates may be artificially inflated by multidimensionality.