

Assessing measurement invariance of a health-related quality-of-life questionnaire in radiotherapy patients

Bellinda L. King-Kallimanis · Claartje L. ter Hoeven ·
Hanneke C. de Haes · Ellen M. Smets ·
Caro C. E. Koning · Frans J. Oort

Accepted: 9 December 2011 / Published online: 23 December 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract

Objective If the assumption of measurement invariance is not tested, we cannot be sure whether differences observed are due to true differences in health-related quality-of-life (HRQoL), or are measurement artifacts. We aim to investigate this assumption in a sample of heterogeneous cancer patients, focusing on whether age, sex, previous treatment for cancer, and information regarding treatment preferences result in biased HRQoL scores.

Methods 155 cancer patients who were about to begin their first session of radiotherapy were included. HRQoL was measured using the EORTC QLQ-C30. Structural equation modeling was applied to assess whether there was a violation of the assumption of invariance.

Results A satisfactory single construct (Functioning HRQoL) measurement model was found and two violations of invariance were identified. Irrespective of patients' Functioning HRQoL, older patients reported worse physical functioning and patients who had received treatment prior to radiotherapy reported worse emotional functioning than we would otherwise expect.

Conclusions In the present study, accounting for measurement bias lead to a substantial improvement in the overall fit of the model. By ignoring the bias, we would have concluded that the model fit was unsatisfactory. The findings underline the importance of investigating measurement invariance in scales designed for heterogeneous samples.

Keywords Measurement invariance · Restricted factor analysis · Health-related quality-of-life · Radiotherapy

Introduction

Many questionnaires have been developed to assess the different facets of cancer patients' experiences before, during, and after their treatment. For example, there are questionnaires to assess their satisfaction with care, their health-related quality-of-life (HRQoL), or their preferred communication style with their oncologists. When developing scales for a general cancer population or testing differences between groups using well-established questionnaires, an important question to keep in mind is whether members of different groups assign the same meaning to questionnaire items. In other words, if there are two patients with the same level of overall satisfaction, will they respond to an observed item in the same way, or will specific characteristics, like gender or treatment regime, influence their response to the item. If it can be shown that these characteristics do not affect responses to observed items, then the assumption of measurement invariance has been met.

The assumption of measurement invariance requires that the relationships between the observed items and the latent construct remain constant regardless of respondents' group membership, for example, age, race, or disease characteristics or the measurement occasion [1, 2]. If this assumption is violated, then the results from cross-group comparisons of the construct may be incorrect. This is because mean differences should represent true differences in the construct of interest and not reflect anything else. For example, it may be that a male patient and female patient share the same underlying level of Physical HRQoL.

B. L. King-Kallimanis (✉) · C. L. ter Hoeven ·
H. C. de Haes · E. M. Smets · C. C. E. Koning · F. J. Oort
Academic Medical Centre, University of Amsterdam,
Amsterdam, The Netherlands
e-mail: kingkalb@tcd.ie

However, when asked a question about carrying groceries, the male who does not do the shopping may respond that he has no difficulty with this activity, whereas the female may indicate that she has great difficulty. The responses given to this grocery item are related not only to Physical HRQoL but also to gender roles. In this example, it is clear to see how gender roles and Physical HRQoL can become entangled. However, it may not always be obvious how patient characteristics might affect certain items. In a study by Reker and Fry [3], bias with respect to age was found in personal meaning measures. The authors concluded that bias in the Self-Transcendence Scales stemmed from older adults using events from the past as their frame of reference, whereas younger adults used present and future events as their frame of reference. When developing items for a scale, this type of bias will be difficult to anticipate and success can only be evaluated after scale development and piloting. If invariance testing yields positive results, in that the measurement is invariant, we can be confident that our results are not distorted because of different functioning of the measurement as a result of group membership. Unfortunately, measurement invariance of self-report questionnaires is often not investigated.

Establishing that a scale has good reliability and validity does not ensure that the scale will not violate the assumption of invariance. The European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30, a measure developed to assess HRQoL, is considered to have excellent reliability and validity [4]. Specifically, it is a generic HRQoL measure for use with all cancer patients, with additional modules for specific cancer diagnosis (e.g., breast [5] and lung cancer [6]). Despite having excellent reliability and validity, the factor structure of this scale has received little attention [7, 8], and most measurement invariance testing has been conducted using item response theory (IRT) [9], with the primary focus on language translation [10, 11]. While designed for a general and therefore heterogeneous cancer population, it is possible that this heterogeneity will lead to a violation of measurement invariance. If we are interested in, for example, differences in HRQoL based on different treatment stages or information preferences, then before differences can be investigated, we must check whether measurement bias with respect to these variables is present. For example, patients who have already received treatment for their diagnosis may have experienced an unmeasured response shift [12]. This in turn could result in a shift in internal standards when responding to HRQoL items, whereas yet to be treated patients will not have experienced this phenomenon. In regard to patients' information preferences, it is conceivable that patients with high compared to low levels of information preferences may respond to HRQoL items using a different frame of reference. This might be

because patients who want more information may want this information to inform their family and friends of their treatment and prognosis [13], therefore, they might have a different frame of reference toward social functioning. Thus, before we investigate the relationships between these variables and HRQoL, we need to be sure that differences in HRQoL mean the same thing for patients in different treatment stages or with different information preferences.

Testing the assumption of measurement invariance in different situations and with different groups of people has been greatly facilitated by the development of several analytic techniques including IRT and structural equation modeling (SEM)/confirmatory factor analysis (CFA) [14]. Within the framework of SEM, there are three approaches available to assess whether the assumption of invariance holds. In cross-sectional research, multi-group CFA comparisons are the most frequently used method. However, to conduct such an analysis, a large sample is required as the sample must be split by group membership. Also, if the potential violator of invariance is continuous, the variable must be transformed to a discrete variable to create multiple groups, in doing this, there is a loss of information. Restricted factor analysis (RFA) is one alternative. The RFA specification allows for multiple groups to be tested simultaneously (i.e., sex and race) and continuous variables can be included as originally measured (i.e., age). These additional variables are modeled as single indicator exogenous variables in the RFA model and tested as possible violators of invariance [15, 16]. The RFA model is equivalent in overall fit and yields the same results as the third alternative, the multiple indicator, multiple cause (MIMIC) model. The difference between these two models is in how the relationships between the exogenous variables and the common factor(s) are modeled. In the MIMIC model, these relationships are causal and in RFA the relationships are not necessarily causal [17]. As we do not necessarily expect causal relationships between the exogenous variables and HRQoL and RFA has been shown in simulation studies to be a robust method [15, 16], we will use RFA.

By using the RFA approach, we can obtain further insight into the psychometric properties of the EORTC QLQ-C30 in a heterogeneous cancer sample. To achieve this we include and study simultaneously multiple variables that have the potential to violate the assumption of measurement invariance. Therefore, the aim of this paper is to investigate whether HRQoL scales are invariant with respect to age, sex, previous treatment for cancer, and patients' information preferences. If any of the observed scales are biased with respect to the exogenous variables, group comparisons in relation to the variables investigated will be less meaningful. So, in doing this, we aim to better understand the construct of HRQoL as measured in a heterogeneous cancer population.

Methods

Participants and procedure

The current study constitutes a part of a larger research project, involving the use of several questionnaires as well as videotaping of the patients' initial- and first follow-up consultation with the radiotherapist. Fifteen radiation oncologists of the radiotherapy department of the Academic Medical Centre in Amsterdam, the Netherlands, were invited to participate in the project. All agreed. Their consecutive newly registered patients were contacted by mail inviting them to participate in the study. Exclusion criteria were: (1) having undergone radiotherapy treatment before; (2) age < 18; (3) unable to understand the Dutch language; and (4) suffering cognitive limitation or cerebral malignancy. Patients were asked for written informed consent, and were invited to fill out a questionnaire at home, prior to their first visit to a radiation oncologist. Non-responding patients were asked to fill in some background variables and one item measuring overall information need. The study was approved by the hospital's medical ethics committee. For further study details, see [18, 19].

Measures

HRQoL

The Dutch language version of the EORTC QLQ-C30 was used to measure HRQoL [4]. It consists of 30 items, 15 items are used to create five subscales related to functioning, which include, Physical Functioning (5 items), Role Functioning (2 items), Emotional Functioning (4 items), Cognitive Functioning (2 items), and Social Functioning (2 items). Two items are used to measure Global Health Status. Thirteen items relate to symptoms experienced by the patient, seven of which are used to create three subscales, Fatigue (3 items), Nausea and Vomiting (2 items), and Pain (2 items). The remaining six items are single-item symptom scales. In this paper we focus only on the multi-item scales and not the single-item scales. Higher scores indicate better HRQoL in regard to functioning; higher scores indicate worse HRQoL in regard to symptoms.

Information preference

To measure a general level of information preference, we used one item from the Information Styles Questionnaire [20]. This item asked patients to indicate their information preference concerning disease and treatment on a 10-point response scale, ranging from 0 (*no information needed*) to 10 (*prefers to receive all available information*).

Previous treatment

Patients' medical records were examined to identify whether they had received either chemotherapy or surgery to treat the same cancer tumor that was being treated by radiotherapy. This information was dichotomized; no treatment compared to previous treatment (chemotherapy/surgery).

Patient characteristics

We considered patients' gender (0 = *male*; 1 = *female*) and their age (continuous).

Statistical analysis

To investigate measurement invariance, we used a two-step procedure; Step 1 involved establishing a measurement model using CFA and Step 2 tested the assumption of invariance with respect to specific patient variables (exogenous variables) by extending the CFA and using RFA. Maximum likelihood estimation was used and all analyses were conducted using the computer program Mx 3.2 [21].

Step 1: Establishing a measurement model

As there is no agreed upon CFA structure for the EORTC QLQ-C30, we aimed to find a satisfactory measurement model for the EORTC QLQ-C30 scales using CFA in our sample. For simplicity we focused only on the multi-item scales and investigate measurement invariance at the scale level. While the use of the symptom scales is controversial, we aimed to include them in our measurement model. Therefore, we fit three measurement models, two of which included the symptom scales, and a third that focused solely on the functioning scales. In Model 1.1, all nine scales loaded on one general HRQoL common factor, and in Model 1.2, all nine scales loaded on two common factors; Functioning HRQoL, which included the five functioning scales and the Global Health Status scale, and, Symptom HRQoL, which included the three symptom scales. Finally, in Model 1.3, the symptom scales were removed, and the five functioning scales and the Global Health Status scale loaded on one common factor, Functioning HRQoL.

To assess the overall goodness-of-fit of our models, the Chi-square test of exact fit, the root mean square error of approximation (RMSEA), and expected cross-validation index (ECVI) were considered [22]. A non-significant Chi-square value indicates good fit; however, it is sensitive to small deviations between the model and data. Therefore, we also considered the RMSEA and ECVI. An RMSEA

value of <0.08 indicates satisfactory fit and a value of <0.05 indicates close fit [22]. The ECVI is used to assess the fit of nested alternative models, in other words it cannot be used as a stand-alone index; smaller values indicate improved model fit [22]. In addition to these overall model fit statistics, we also considered the standardized residuals to identify potential sources of model misfit and if required, guide appropriate model modifications.

If a new model was specified, we investigated the change in overall model fit by using both the Chi-square difference test and ECVI difference test. The Chi-square difference test is the difference in Chi-square values between the alternative and null models; if the difference is significant, the re-specification has improved model fit. The ECVI difference test is the difference in ECVI values for the alternative and null models; if the 90% CI does not include zero, then the re-specification has improved model fit [22]. It complements the Chi-square difference test, but it penalizes models containing more free parameters.

Step 2: Testing invariance with respect to exogenous variables

Using the final model from Step 1, we included all additional exogenous variables in the model. These included age, sex, previous treatment, and information preferences. These additional variables were allowed to correlate with the latent variable(s), but all direct effects of these variables on the observed scales were fixed to zero. A violation of invariance is indicated by a significant direct effect of an exogenous variable on an observed variable.

In order to identify significant direct effects, a series of iterative tests were conducted. We fit models where the direct effect between the exogenous variable and the observed scale under consideration was freed. For example, when investigating invariance associated with sex, we fit a series of models with an additional parameter for the effect of sex on each one of the EORTC QLQ-C30 scales included in the final measurement model from Step 1. This resulted in a series of one degree of freedom Chi-square difference tests. If any of these tests were significant at a Bonferroni corrected significance level [23] and the observed parameter change (OPC) was greater than 0.1, then we considered the scale to be non-invariant in relation to sex. The OPCs are the difference between the standardized parameter in the null model (equal to zero in this example) and the standardized parameter estimated in the test. We rely on Chi-square difference tests and OPCs rather than modification indices and expected parameter change because particularly the modification indices can be influenced by mis-specification elsewhere in the model [24]. We used a cut-point for the OPC of 0.1, which was based on Cohen's small effect sizes [25].

Results

Sample

The questionnaire was sent to 293 eligible radiotherapy patients of whom 159 (54%) agreed to participate. Non-responding patients were on average older than the participating patients ($M = 66.61$, $SD = 13.49$ versus $M = 62.98$, $SD = 12.64$; $P = 0.019$); no gender differences were found. In this paper we only focus on assessment before their initial radiotherapy consultation. Four patients had missing values on treatment information and were therefore excluded from further analyses, leaving 155 participants in the sample. Characteristics of the participating patients are given in Table 1. In Table 2 we present the correlations, means, and standard deviations of the items included in the RFA analysis. As can be seen, the assumption of multivariate normality was violated, therefore the resulting test statistic may not have a central Chi-square distribution, and the standard errors may not be correct [26]. Hoogland and Boomsma [27] suggest that this does not seriously bias the estimates of the model parameters in samples larger than 200; however, little is known about the effects of violations of multivariate normality in smaller samples such as the present sample.

Step 1: Establishing a measurement model

We tested our three possible measurement models, and found all three models to have unsatisfactory overall model fit (Model 1.1; 1.2; 1.3 Table 3). Suggested modifications via the standardized residuals for Models 1.1 and 1.2 were

Table 1 Descriptive statistics for demographic variables of radiotherapy patients ($N = 155$)

Variable	Number (%)
Sex—female	60 (38.71%)
Age (mean and SD)	62.98 (12.64)
Previous treatment—yes	65 (41.94%)
Information preference (mean and SD)	8.68 (1.98)
<i>Cancer site</i>	
Prostate	33 (21.43%)
Breast	24 (15.58%)
Rectal/colorectal	19 (12.34%)
Esophageal	16 (10.39%)
Lung	13 (8.44%)
Endometrial	7 (4.55%)
Cervical	6 (3.90%)
Bladder	6 (3.90%)
Other	31 (20.00%)

Other cancer; e.g., gallbladder, testicular, pancreas, and non-Hodgkin lymphoma

Table 2 Correlations, means, and standard deviations for all observed variables

	PF	RF	EF	CF	SF	GH	Age	Sex	Prev. tx	Info. pref.
PF	1.00									
RF	0.61***	1.00								
EF	0.18*	0.36***	1.00							
CF	0.52***	0.49***	0.32***	1.00						
SF	0.59***	0.69***	0.36***	0.41***	1.00					
GH	0.54***	0.60***	0.43***	0.39***	0.54***	1.00				
Age	-0.16*	0.22**	0.16*	0.02	0.12	0.03	1.00			
Sex	-0.18*	-0.14	-0.13	-0.06	-0.16*	-0.10	-0.25**	1.00		
Prev. tx	0.22**	-0.01	-0.20*	0.13	-0.003	0.03	-0.41***	0.24**	1.00	
Info pref	-0.03	-0.07	-0.01	-0.11	-0.03	0.15	-0.10	0.05	0.04	1.00
Mean (SD)	7.97 (2.05)	6.97 (3.11)	7.45 (1.94)	8.61 (1.86)	8.09 (2.36)	7.09 (2.17)	62.98 (12.64)	0.39 (0.49)	0.42 (0.49)	8.56 (2.26)

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$

difficult to interpret. Therefore, no modifications were made to these models. Model 1.3 included one source of misfit, a residual covariance between Physical Functioning and Emotional Functioning. With the inclusion of this additional parameter the model had satisfactory overall fit (Model 1.F, Table 3; Fig. 1) and all factor loadings were significant.

Step 2: Testing invariance with respect to exogenous variables

We added the exogenous variables to the final model of Step 1 (Model 2, Table 3). The overall fit of the new model was not satisfactory. Although the parameter estimates of this model cannot be trusted due to poor fit, we still looked at the strength of the correlations between the exogenous variables and Functioning HRQoL before investigating invariance. The largest correlation observed in this model was between sex (-0.20) and Functioning HRQoL. The correlations between information preferences (-0.09), age (0.09), and previous treatment (0.05) were considered to be small.

In the series of tests investigating invariance, two violations of invariance were identified as the OPC was greater than 0.1, and both the Chi-square difference test and ECVI difference tests were significant. The first significant direct effect was with age on Physical Functioning (Model 2.1). The next and last significant direct effect identified was with previous treatment on Emotional Functioning (Model 2.F). After this iteration, the next largest direct effect was of information preferences on Global Health status (Model 2.3). The OPC was greater than 0.1 (-0.15); however, both the Chi-square difference test (according to Bonferroni adjustment) and ECVI difference test were not significant, therefore this direct effect was not included in the final model.

After the inclusion of these two direct effects, the overall fit of the model was satisfactory (Model 2.F, final parameter estimates Fig. 1). We re-examined the correlations between the exogenous variables and Functioning HRQoL to investigate the impact the inclusion of the additional direct effects had on the correlations. The correlations between age, gender, and information preferences and Functioning HRQoL increased. The largest increase was between age and Functioning HRQoL, the correlation increased from 0.09 to 0.17. There were slight increases in the correlations between gender and Functioning HRQoL (-0.20 to -0.22) and between information preferences and Functioning HRQoL (-0.09 to -0.10). There was no change in the correlation between previous treatment and Functioning HRQoL (0.05). See Fig. 1 for all final model parameter estimates.

We can now interpret how the two direct effects impact the observed scales (Physical Functioning and Emotional Functioning) in relation to the correlations between age and previous treatment and Functioning HRQoL. The direct effect of age on Physical Functioning was negative (-0.30) and suggests that older patients report worse PF than would be expected given there is a weak positive correlation between age and Functioning HRQoL. The direct effect of previous treatment on Emotional Functioning was also negative (-0.20) and suggests patients who have previously received treatment related to their current diagnosis report worse Emotional Functioning than would be expected given there is almost no relationship between previous treatment and Functioning HRQoL.

Discussion

In this study, we investigated the assumption of measurement invariance for the EORTC QLQ-C30 in a heterogeneous population of cancer patients who were about to

Table 3 Overall goodness-of-fit and Chi-square difference test results for EORTC QLQ-C30

Model	CHISQ (df)	P value	RMSEA (90% CI)	ECVI (90% CI)	Comparison models	CHISQ DIFF (df)	P value	ECVI DIFF (90% CI)
1.1 Measurement model 1	58.0 (27)	<0.001	0.086 (0.056; 0.120)	0.610 (0.499; 0.780)				
1.2 Measurement model 2	57.74 (26)	<0.001	0.089 (0.058; 0.120)	0.621 (0.501; 0.790)				
1.3 Measurement model 3	28.1 (9)	0.001	0.100 (0.053; 0.150)	0.317 (0.247; 0.440)				
1.F Final measurement model	14.7 (8)	0.064	0.072 (0.000; 0.131)	0.270 (0.231; 0.368)	1.F versus 1.3	13.3 (1)	0.0003	0.073 (0.018; 0.178)
2 Addition of exogenous variables	76.9 (28)	<0.001	0.107 (0.078; 0.135)	0.877 (0.729; 1.079)				
2.1 Age and physical functioning	50.3 (27)	0.004	0.075 (0.041; 0.107)	0.718 (0.612; 0.879)	2.1 versus 2	26.6 (1)	<0.001	0.159 (0.070; 0.299)
2.F Previous treatment and emotional functioning	42.3 (26)	0.023	0.064 (0.024; 0.098)	0.680 (0.589; 0.827)	2.F versus 2.1	8.0 (1)	0.005	0.038 (0.001; 0.127)
2.3 Information preferences and global health status	36.1 (25)	0.070	0.054 (0.000; 0.090)	0.654 (0.594; 0.778)	2.F versus 2.3	6.2 (1)	0.013	0.026 (-0.003; 0.1074)

df degrees of freedom, *RMSEA* root mean square error of approximation, *ECVI* expected cross-validation index

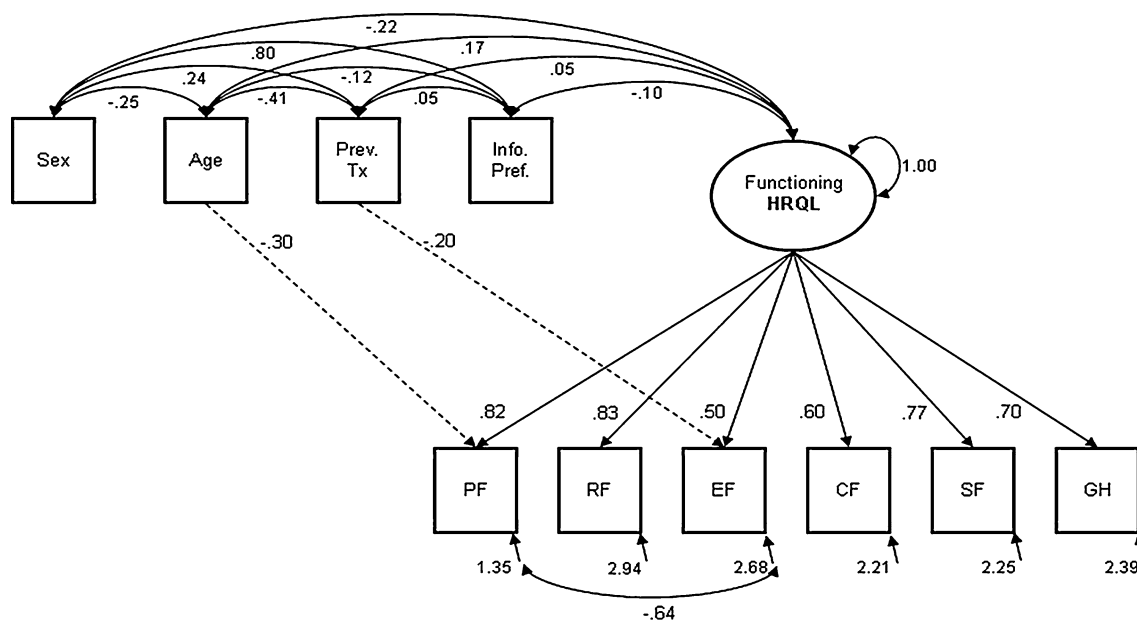


Fig. 1 EORTC QLQ-C30 restricted factor analysis model: standardized parameter estimates from model 2.F. *Note:* PF physical functioning, RF role functioning, EF emotional functioning, CF cognitive functioning, SF social functioning, GH global health status

begin their first session of radiotherapy. Applying RFA, we investigated age, sex, previous treatment for cancer, and information preferences regarding treatment as potential violators of invariance. Two violations were identified in Physical Functioning with regard to age and in Emotional Functioning with regard to previous treatment.

In the first step, we were able to fit a measurement model to the EORTC QLQ-C30 that had satisfactory fit. Our final measurement model did not include any of the symptoms scales of the EORTC QLQ-C30. However, Fayers and Hand have argued that these symptom scales should not be used as manifestations of underlying HRQoL but rather as manifestations of treatment [28]. This is because one would expect a different factor structure for symptoms dependent on the type of treatment the patient was undergoing. While this substantive debate is beyond the scope of this paper, in the current sample, the patients are in different stages of treatment, and this may explain why the inclusion of symptom scales did not lead to a satisfactory model. Once we identified a satisfactory measurement model we were able to investigate invariance in the EORTC QLQ-C30, therefore, it was in Step 2 that we identified the two violations of invariance.

The direct effect between age and Physical Functioning suggested that if younger and older patients had the same underlying Functioning HRQoL, older patients reported their Physical Functioning to be worse than younger patients. This result was found in another study where measurement invariance was investigated in regard to the

SF-36 [29] (HRQoL measure) in a sample of cancer patients [30]. The authors suggested that because Physical Functioning is the most objective HRQoL scale, leaving little room for individualized interpretation, it is conceivable that it is the other scales that are biased because there is more room for subjective interpretation. An alternative model could be fitted that allowed direct effects between age and the other scales, excluding Physical Functioning; however, this model would include many additional parameters and as a result be less parsimonious. Therefore, we opt for parsimony and the model with least instances of measurement bias. As a result of our finding, care should be taken when making age comparisons with any of the EORTC QLQ-C30 scales and age. Recently, the EORTC QLQ-ELD15 [31] was developed specifically for older adults, though it is ideal to have observed variables that are invariant to the effects of age.

The direct effect between previous treatment and Emotional Functioning was also negative. In other words, radiotherapy patients who had undergone a previous treatment (chemotherapy/surgery) evaluated their Emotional Functioning worse than those who did not undergo treatment before starting radiotherapy, even when their underlying Functioning HRQoL was similar. The different interpretation of Emotional Functioning might be due to resource depletion [32]. According to resource models, self-regulatory resources can be depleted or fatigued by self-regulatory demands. Muraven et al. [33] found that one route to self-regulatory failure is prior self-regulatory activities. In their laboratory studies, participants who were

asked to employ a form of self-regulation (e.g., mental control or regulation of emotional expression) were less able to self-regulate after that (see also [34]). Previous treatment can be regarded as a prior self-regulatory effort, where emotions needed to be regulated. To undergo more treatment might decrease Emotional Functioning, because regulatory resources are depleted. This depletion may result in a different frame of reference in regard to Emotional Functioning for patients who have already undergone treatment. Interestingly, the latent construct of Functioning HRQoL was not reliant on self-regulatory efforts as evidenced by the very small relationship between previous treatment and Functioning HRQoL. To better understand this relationship, more research with longitudinal data is needed.

Previous research has shown that the EORTC QLQ-C30 has excellent psychometric properties [35] and is used extensively to assess HRQoL [36–38]. For the aim of our study to investigate invariance, we believe the model we used was a good representation of the functioning scales of the EORTC QLQ-C30. The two instances of non-invariance detected do not suggest that Physical Functioning and Emotional Functioning are not valid indicators of Functioning HRQoL, but rather that care should be taken when using the functioning scales to compare younger and older adults and patients at different stages of treatment. While our sample size was small, and therefore limits generalization, the direct effect between age and Physical Functioning has been identified in previous research, indicating that it is certainly worth further investigation in a longitudinal study. In addition to this, it would be worthwhile to also consider invariance of the EORTC QLQ-C30 with respect to different cancer diagnoses, different treatment regimes, and different stages of cancer. Focusing on these specific variables would lead to greater confidence when comparing differences in HRQoL in relation to these variables.

Accounting for violations to the assumption of measurement invariance in our study lead to a significant improvement in overall model fit. The inclusion of patient characteristic variables to our model initially resulted in a model where the estimates could not be confidently interpreted. However, after the inclusion of direct effects accounting for bias, our model fit was satisfactory and conclusions regarding the model could be drawn. It is important to note that a single violation of invariance may not be enough to cause unsatisfactory model fit, but could have a substantial impact on the conclusions drawn. In other words, if the assumption of measurement invariance is ignored the researcher cannot be sure whether differences observed are related to true differences in HRQoL, or whether these differences are related to how patients interpret the HRQoL items.

Acknowledgments The Dutch Cancer Society funded this study n (grant UVA 2005-3199). We like to thank the patients who participated in this study.

Conflict of interest The authors of this paper have no conflicts of interests to report that may bias the findings of this study.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, 13(2), 127–143.
- Meredith, W. (1993). Measurement invariance, factor-analysis and factorial invariance. *Psychometrika*, 58(4), 525–543.
- Reker, G. T., & Fry, P. S. (2003). Factor structure and invariance of personal meaning measures in cohorts of younger and older adults. *Personality and Individual Differences*, 35(5), 977–993.
- Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., et al. (1993). The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute*, 85(5), 365–376.
- Sprangers, M. A. G., Groenvold, M., Arraras, J. I., Franklin, J., te Velde, A., Muller, M., et al. (1996). The European Organization for Research and Treatment of Cancer breast cancer-specific quality-of-life questionnaire module: First results from a three-country field study. *Journal of Clinical Oncology*, 14(10), 2756–2768.
- Bergman, B., Aaronson, N. K., Ahmedzai, S., Kaasa, S., & Sullivan, M. (1994). The Eortc Qlq-Lc13—A modular supplement to the EORTC core quality-of-life questionnaire (Qlq-C30) for use in lung-cancer clinical-trials. *European Journal of Cancer*, 30A(5), 635–642.
- Gotay, C. C., Blaine, D., Haynes, S. N., Holup, J., & Pagano, I. S. (2002). Assessment of quality of life in a multicultural cancer patient population. *Psychological Assessment*, 14(4), 439–450.
- Snyder, C. F., Garrett-Mayer, E., Brahmer, J. R., Carducci, M. A., Pili, R., Stearns, V., et al. (2008). Symptoms, supportive care needs, and function in cancer patients: How are they related? *Quality of Life Research*, 17(5), 665–677.
- Scott, N. W., Fayers, P. M., Aaronson, N. K., Bottomley, A., de Graeff, A., Groenvold, M., et al. (2010). Differential item functioning (DIF) analyses of health-related quality of life instruments using logistic regression. *Health and Quality of Life Outcomes*, 8, 81.
- Petersen, M. A., Groenvold, M., Bjorner, J. B., Aaronson, N., Conroy, T., Cull, A., et al. (2003). Use of differential item functioning analysis to assess the equivalence of translations of a questionnaire. *Quality of Life Research*, 12(4), 373–385.
- Scott, N. W., Fayers, P. M., Bottomley, A., Aaronson, N. K., de Graeff, A., Groenvold, M., et al. (2006). Comparing translations of the EORTC QLQ-C30 using differential item functioning analyses. *Quality of Life Research*, 15(6), 1103–1115.
- Sprangers, M. A. G., & Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: A theoretical model. *Social Science and Medicine*, 48(11), 1507–1515.
- Rees, C. E., & Bath, P. A. (2000). The information needs and source preferences of women with breast cancer and their family members: A review of the literature published between 1988 and 1998. *Journal of Advanced Nursing*, 31(4), 833–841.

14. Stark, S., Chernyshenko, E. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology, 91*(6), 1292–1306.
15. Oort, F. J. (1998). Simulation study of item bias detection with restricted factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 5*(2), 107–124.
16. Barendse, M. T., Oort, F. J., & Garst, G. J. A. (2010). Using restricted factor analysis with latent moderated structures to detect uniform and nonuniform measurement bias; a simulation study. *Asta-Advances in Statistical Analysis, 94*, 117–127.
17. Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54*(4), 557–585.
18. ter Hoeven, C. L., Zandbelt, L. C., Franssen, S., de Haes, H., Oort, F., Geijsen, D., Koning, C., & Smets, E. (2010). Measuring cancer patients' reasons for their information preference: Construction of the considerations concerning cancer information (CCCI) questionnaire. *Psycho-Oncology*. doi:10.1002/pon.184.
19. Zeguers, M., de Haes, H. C., Zandbelt, L. C., ter Hoeven, C. L., Franssen, S. J., Geijsen, D. D., Koning, C. C., & Smets, E. M. (2010). The information needs of new radiotherapy patients: How to measure? Do they want to know everything? And if not, why? *International Journal of Radiation Oncology Biology Physics*. doi:10.1016/j.ijrobp.2010.09.032.
20. Cassileth, B. R., Zupkis, R. V., Sutton-Smith, K., & March, V. (1980). Information and participation preferences among cancer patients. *Annals of Internal Medicine, 92*(6), 832–836.
21. *MxGui*. (2010). [computer program]. Version 32. Richmond, VA: VCU.
22. Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*(2), 230–258.
23. Tabachnick, B. G., & Fidell, L. S. (2006). *Using multivariate statistics* (5th ed.). Needham Heights, MA, USA: Allyn & Bacon, Inc.
24. Saris, W. E., Satorra, A., & Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. *Sociological Methodology, 17*, 105–129.
25. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum associates.
26. Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
27. Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling—An overview and a meta-analysis. *Sociological Methods & Research, 26*, 329–367.
28. Fayers, P. M., & Hand, D. J. (1997). Factor analysis, causal indicators and quality of life. *Quality of Life Research, 6*(2), 139–150.
29. Ware, J. E., Snow, K. K., Kosinski, M., & Gandek, B. (1993). *SF-36 health survey manual and interpretation guide*. Boston, MA: The Health Institute, New England Medical Center.
30. King-Kallimanis, B. L., Oort, F. J., Visser, M. R., & Sprangers, M. A. G. (2009). Structural equation modeling of health-related quality-of-life data illustrates the measurement and conceptual perspectives on response shift. *Journal of Clinical Epidemiology, 62*(11), 1157–1164.
31. Johnson, C., Fitzsimmons, D., Gilbert, J., Arraras, J. I., Hammerlid, E., Bredart, A., et al. (2010). Development of the European Organisation for Research and Treatment of Cancer quality of life questionnaire module for older people with cancer: The EORTC QLQ-ELD15. *European Journal of Cancer, 46*(12), 2242–2252.
32. Vohs, K. D., & Heatherton, T. F. (2000). Self-regulatory failure: A resource-depletion approach. *Psychological Science, 11*(3), 249–254.
33. Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-control as limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology, 74*(3), 774–789.
34. Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology, 74*(5), 1252–1265.
35. Kaasa, S., Bjordal, K., Aaronson, N., Moum, T., Wist, E., Hagen, S., et al. (1995). The EORTC core quality of life questionnaire (QLQ-C30): Validity and reliability when analysed with patients treated with palliative radiotherapy. *European Journal of Cancer, 31A*(13–14), 2260–2263.
36. Slotman, B. J., Mauer, M. E., Bottomley, A., Faivre-Finn, C., Kramer, G. W. P. M., Rankin, E. M., et al. (2009). Prophylactic cranial irradiation in extensive disease small-cell lung cancer: Short-term health-related quality of life and patient reported symptoms—results of an international phase III randomized controlled trial by the EORTC radiation oncology and lung cancer groups. *Journal of Clinical Oncology, 27*(1), 78–84.
37. Mosconi, P., Apolone, G., Barni, S., Secondino, S., Sbanotto, A., & Filiberti, A. (2002). Quality of life in breast and colon cancer long-term survivors: An assessment with the EORTC QLQ-C30 and SF-36 questionnaires. *Tumori, 88*(2), 110–116.
38. Greimel, E. R., Winter, R., Kapp, K. S., & Haas, J. (2009). Quality of life and sexual functioning after cervical cancer treatment: A long-term follow-up study. *Psycho-Oncology, 18*(5), 476–482.