



UvA-DARE (Digital Academic Repository)

Taking into account the impact of attrition on the assessment of response shift and true change: a multigroup structural equation modeling approach

Verdam, M.G.E.; Oort, F.J.; van der Linden, Y.M.; Sprangers, M.A.G.

DOI

[10.1007/s11136-014-0829-y](https://doi.org/10.1007/s11136-014-0829-y)

Publication date

2015

Document Version

Final published version

Published in

Quality of Life Research

[Link to publication](#)

Citation for published version (APA):

Verdam, M. G. E., Oort, F. J., van der Linden, Y. M., & Sprangers, M. A. G. (2015). Taking into account the impact of attrition on the assessment of response shift and true change: a multigroup structural equation modeling approach. *Quality of Life Research*, 24(3), 541-551. <https://doi.org/10.1007/s11136-014-0829-y>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

Taking into account the impact of attrition on the assessment of response shift and true change: a multigroup structural equation modeling approach

Mathilde G. E. Verdam · Frans J. Oort ·
Yvette M. van der Linden · Mirjam A. G. Sprangers

Accepted: 10 October 2014 / Published online: 18 October 2014
© Springer International Publishing Switzerland 2014

Abstract

Purpose Missing data due to attrition present a challenge for the assessment and interpretation of change and response shift in HRQL outcomes. The objective was to handle such missingness and to assess response shift and ‘true change’ with the use of an attrition-based multigroup structural equation modeling (SEM) approach.

Method Functional limitations and health impairments were measured in 1,157 cancer patients, who were treated with palliative radiotherapy for painful bone metastases, before [time (T) 0], every week after treatment (T1 through T12), and then monthly for up to 2 years (T13 through T24). To handle missing data due to attrition, the SEM procedure was extended to a multigroup approach, in which we distinguished three groups: short survival (3–5 measurements), medium survival (6–12 measurements), and long survival (>12 measurements).

Results Attrition after third, sixth, and 13th measurement occasions was 11, 24, and 41 %, respectively. Results show that patterns of change in functional limitations and health impairments differ between patients with short, medium, or long survival. Moreover, three response-shift effects were detected: recalibration of ‘pain’ and ‘sickness’ and

reprioritization of ‘physical functioning.’ If response-shift effects would not have been taken into account, functional limitations and health impairments would generally be underestimated across measurements.

Conclusions The multigroup SEM approach enables the analysis of data from patients with different patterns of missing data due to attrition. This approach does not only allow for detection of response shift and assessment of true change across measurements, but also allow for detection of differences in response shift and true change across groups of patients with different attrition rates.

Keywords Structural equation modeling · Response shift · Health-related quality of life · Attrition · Missing data · Multigroup

Introduction

Missing data due to attrition present a challenge for the assessment and interpretation of change in HRQL outcomes, as it is often related to a declining health status [1]. In particular, in clinical settings, missing data due to attrition are likely to be missing not at random (MNAR), i.e., when attrition is related to the values of HRQL that are missing. MNAR thus poses a problem for data analysis. Data imputation methods might lead to biased estimates of change under the condition of MNAR and cannot be sensibly applied when attrition is related to patients’ declining health status. One option is to only analyze data from patients who were able to complete all measurements, i.e., complete case analysis. Another option is to only use data from those measurements that all patients completed. However, in both situations, important information is lost and results can be misleading. For example, patients who

M. G. E. Verdam (✉) · F. J. Oort · M. A. G. Sprangers
Department of Medical Psychology, Academic Medical Centre,
University of Amsterdam, Amsterdam, The Netherlands
e-mail: m.g.e.verdam@uva.nl

M. G. E. Verdam · F. J. Oort
Department of Child Development and Education, University of
Amsterdam, Amsterdam, The Netherlands

Y. M. van der Linden
Department of Clinical Oncology, Leiden University Medical
Centre, Leiden, The Netherlands

drop out early in a clinical trial may have more severe disease trajectory than patients who drop out at a later stage in the trial. It is therefore important to account for attrition when analyzing data from longitudinal clinical trials.

Attrition may be related to changes in HRQL and thus also to ‘response shifts.’ Response shift refers to a change in the meaning of one’s self-evaluation of a target construct that may cause changes in observed variables (e.g., responses to questionnaires) that are not directly related to change in the construct of interest (e.g., HRQL). Sprangers and Schwartz [2] distinguish three types of response shift: (1) recalibration, which refers to a change in the respondent’s internal standards of measurement; (2) reprioritization, which refers to a change in respondent’s values regarding the relative importance of component domains of the target construct; and (3) reconceptualization, which refers to a change in definition of the target construct. When we assess change in patient-reported HRQL outcomes, it is important to also investigate—and account for—response-shift effects.

Structural equation modeling (SEM) can be used to detect various types of response shift, to account for them and to measure ‘true change’ [3]. True change refers to a change in the patient’s level of the target construct (e.g., an improvement or deterioration of HRQL), while response shift refers to other changes that may obfuscate true changes (e.g., recalibration, reprioritization, and reconceptualization response shifts). Each type of response shift can be linked to changes in specific parameter estimates of the model, where changes in the pattern of factor loadings indicate reconceptualization, changes in the values of factor loadings indicate reprioritization, and changes in the values of intercepts indicate (uniform) recalibration.

The aim of this article is to account for attrition in the investigation of changes in HRQL and response-shift effects. We applied the SEM approach [3] to data from patients with painful bone metastases who received palliative treatment. To take into account the possible impact of attrition on the interpretation of findings, the SEM approach was extended to a multigroup approach, in which groups were distinguished based on their pattern of missing values.

Method

Patients

In the Dutch Bone Metastasis Study (DBMS) [4, 5], a total of 1,157 patients (533 women) with painful bone metastases from a solid tumor were enrolled from 17 radiotherapy institutes in the Netherlands. Patients’ primary tumor

was either breast cancer ($n = 451$), prostate cancer ($n = 267$), lung cancer ($n = 287$), or other ($n = 152$). Patients were randomized to receive treatment of a single fraction (8 Gray) versus multiple fractions (six times 4 Gray) of radiation. Possible side effects from radiation therapy vary depending on the part of the body being treated and may include skin changes (dryness, itching, peeling, or blistering), fatigue, loss of appetite, hair loss, diarrhea, nausea, and vomiting. Most of these side effects go away within a few weeks after radiation therapy is finished.

Before treatment (T0) and during the first 12 weeks of follow-up, patients completed weekly HRQL questionnaires by mail (T1 through T12). After that, assessments continued monthly for up to 2 years or until death (T13 through T24). For the present study, we used a subset of data from the DBMS database (T0 through T12) and distinguished three groups based on their pattern of missingness (i.e., attrition rate). Although there is some discrepancy between attrition and actual time of death, we will refer to these groups as: short survival (3–5 measurements; $n = 144$), medium survival (6–12 measurements; $n = 203$), and long survival (>12 measurements; $n = 682$).

Measures

HRQL was assessed with three questionnaires: the Rotterdam Symptom Checklist (RSCL; [6]), the EQ-5D [7], and the EORTC QLQ-C30 [8]. For the present study, questionnaire items were grouped into scales, based on results of principle component analyses. In addition to statistical considerations, we also considered clustering of original questionnaire scales (e.g., the social functioning scale of the EORTC QLQ-C30 [8]) and previous results of factor structure analyses (e.g., the factor structure of the physical symptom distress scale of the RSCL [6]). This resulted in the computation of eight health indicators: physical functioning (PF; 4 items), mobility (MB; 5 items), social functioning (SF; 2 items), depression (DP; 8 items), listlessness (LS; 6 items), pain (PA; 4 items), sickness (SI; 6 items), and treatment-related symptoms (SY; 11 items) (see Table 1). All scale scores were calculated as mean item scores, ranging from 1 to 4, with higher scores indicating more symptoms or more dysfunctioning. Cronbach’s alpha coefficients [9] indicated moderate to good internal consistency reliability (PF, $\alpha = 0.93$; MB, $\alpha = 0.91$; SF, $\alpha = 0.80$; DP, $\alpha = 0.94$; LS, $\alpha = 0.72$; SI, $\alpha = 0.74$; PA $\alpha = 0.74$; SY, $\alpha = 0.69$).

Intermittent missing item scores and missing scale scores were imputed using expectation–maximization [10].

Table 1 Health indicators and allocated questionnaire items used for statistical analyses

Item	Source
<i>Physical functioning</i>	
Light housework/household jobs	RSCL activity level
Heavy housework/household jobs	RSCL activity level
Go shopping	RSCL activity level
Usual activities	EQ-5D
<i>Mobility</i>	
Care for myself	RSCL activity level
Walk about the house	RSCL activity level
Climb stairs	RSCL activity level
Walk out of doors	RSCL activity level
Mobility	EQ-5D
<i>Social functioning</i>	
Limiting social activities	EORTC QLQ-C30 social functioning
Limiting family life	EORTC QLQ-C30 social functioning
<i>Depression</i>	
Anxiety	RSCL psychological distress
Tension	RSCL psychological distress
Worrying	RSCL psychological distress
Nervousness	RSCL psychological distress
Despairing about the future	RSCL psychological distress
Depressed mood	RSCL psychological distress
Irritability	RSCL psychological distress
Anxiety/depression	EQ-5D
<i>Listlessness</i>	
Lack of energy	RSCL physical symptom distress
Tiredness	RSCL physical symptom distress
Difficulty concentrating	RSCL physical symptom distress
Shortness of breath	RSCL physical symptom distress
Difficulty sleeping	RSCL physical symptom distress
Decreased sexual interest	RSCL physical symptom distress
<i>Sickness</i>	
Nausea	RSCL physical symptom distress
Vomiting	RSCL physical symptom distress
Lack of appetite	RSCL physical symptom distress
Acid indigestion	RSCL physical symptom distress
Diarrhea	RSCL physical symptom distress
<i>Pain</i>	
Low back pain	RSCL physical symptom distress
Sore muscles	RSCL physical symptom distress
Pain/discomfort	EQ-5D
Pain in the bones	Developed specifically for DBMS
<i>Treatment-related symptoms</i>	
Painful skin	Developed specifically for DBMS
Itching	Developed specifically for DBMS

Table 1 continued

Item	Source
Sore mouth/pain when swallowing	RSCL physical symptom distress
Dry mouth	RSCL physical symptom distress
Headaches	RSCL physical symptom distress
Burning/sore eyes	RSCL physical symptom distress
Dizziness	RSCL physical symptom distress
Tingling hands or feet	RSCL physical symptom distress
Shivering	RSCL physical symptom distress
Loss of hair	RSCL physical symptom distress
Constipation	RSCL physical symptom distress

RSCL = Rotterdam Symptom Checklist [6]; EORTC QLQ-C30 = European Organization for Research and Treatment of Cancer, Quality of Life Questionnaire C30 [8]; EQ-5D = health outcome instrument of the EuroQol group [7]; DBMS = Dutch Bone Metastasis Study [4, 5]

Per assessment, 27–42 % of respondents showed intermittent missing item scores and 1–8 % of respondents showed intermittent missing scale scores.

Statistical procedure

Structural equation modeling was used to detect response shift and to assess true change [3]. To handle missing data due to attrition, the procedure was extended to a multi-group approach, in which we distinguish three groups based on their pattern of missingness: short survival (3–5 measurements), medium survival (6–12 measurements), and long survival (>12 measurements). This enables to incorporate data from patients with different attrition rates. The SEM procedure included the following steps: (1) establishing an appropriate measurement model, (2) fitting a model of no response shift, (3) detection of response shift, and (4) assessment of true change. These steps are based on the proposed SEM procedure as described by Oort [3], but here, the modeling procedure is modified to enable measurement bias detection in a multigroup model with data from 3, 6, and 13 measurements, respectively.

Step 1: Measurement model

The *Measurement Model* was established on the basis of results of exploratory factor analyses of the present data (exploratory results of the first measurement occasion were confirmed at subsequent measurement occasions) and substantive considerations. The complete longitudinal factor model consists of equivalent factor structures at thirteen consecutive measurement occasions and includes all longitudinal relations between common factors and all longitudinal relations between the same residual factors across

time. To reduce the complexity of the model (i.e., the number of parameter estimates), Kronecker product restrictions were imposed on residual factor variances and covariances to profit from the multivariate longitudinal structure of the data. This restriction entails that the changes in residual factor variances and covariances across occasions are proportionate for all residual factors. The resulting longitudinal three-mode model (L3MM [11]) is more parsimonious and has attractive interpretation. The imposition of these restrictions has previously been illustrated in a subset of the current data [12]. The *Measurement Model* has no equality constraints across occasions or survival groups.

Step 2: No response shift model

To assess the occurrence of response shift, a model of no response shift is fitted to the data, where the model parameters that are associated with response-shift effects (factor loadings and intercepts) are constrained to be equal across measurement occasion and survival group (for ease of presentation, restrictions on residual variances are not considered here). Thus, instead of estimating all factor loadings separately for each measurement occasion and each survival group (9 factor loadings at 3, 6, and 13 measurements, respectively), we now only estimate 9 factor loadings that are constrained to be invariant across measurement occasions and survival groups. Instead of estimating all intercepts separately (8 intercepts at 3, 6, and 13 measurements, respectively), we only estimate 8 invariant intercepts. Equality constraints across survival groups enable the investigation of differences in the occurrence of response shift between patients with different attrition rates. The occurrence of response shift can be assessed by comparing model fit of this restricted model to the model fit of the model with no equality constraints across occasions or groups. When there is no significant deterioration in model fit, the *No Response Shift Model* can be retained.

Step 3: Response shift model

Detection of response shift was done through step-by-step modification of the *No Response Shift Model*, resulting in a model where all apparent response shifts are accounted for. Response shift is operationalized as across-measurement differences between patterns of common factor loadings (reconceptualization), values of common factor loadings (reprioritization), and differences between intercepts (recalibration). Differences across groups may indicate a differential response-shift effect between survival groups. The step-by-step modification of the *No Response Shift Model* was conducted using an iterative procedure where the response-shift parameters associated with reconceptualization, reprioritization, and recalibration were freely

estimated across measurements and groups for all indicators. Thus, in the first iteration, we fitted three models (for each type of response shift) per indicator, resulting in 24 separate models. The model that yielded the largest improvement in model fit was further investigated to determine whether the improvement in fit occurred because of differences across measurements, differences across groups, or both. The response-shift effects and group differences that were found were accounted for by incorporating them in the model.

Step 4: True change

True change was assessed in the model where all apparent response shifts are accounted for. Latent trajectories were inspected by plotting latent factor means across time. Tests of invariance and confidence intervals were used to evaluate differences in common factor means across measurement occasions and between survival groups. To evaluate the impact of response shift on the assessment of change, we inspected the trajectories of common factor means, before and after taking response-shift effects into account.

Statistical analysis

Structural equation models were fitted to the means, variances, and covariances of the eight observed health indicators at three, six, and thirteen measurement occasions (short-, medium-, and long-survival groups, respectively; and 24, 48, and 104 observed variables) using OpenMx [13]. To achieve identification of all model parameters, scales and origins of the common factors were established in Step 1 by fixing the factor means at zero and the factor variances at one. In Step 2, only for first occasion (T0; baseline) of the short-survival group (G1), factor means and variances are fixed; factor means and variances at follow-up occasions (T1–T12) and medium- and long-survival groups (G2–G3) are then identified by constraining intercepts and factor loadings to be equal across assessment and group. Identification of the L3MM-restricted residual variances and covariances was achieved by fixing one element of the matrices that feature in these L3MM restrictions to one (for more details on identification restrictions in the L3MM, the reader is referred to Oort [11]).

To evaluate goodness of fit, the chi-square test of exact fit (CHISQ) was used, where a significant chi-square indicates a significant difference between model and data. However, chi-square values increase with larger sample sizes and more parsimonious models. Alternatively, the root mean square error of approximation (RMSEA [14, 15]) was used, where RMSEA values below 0.05 indicate ‘close’ approximate fit and values below 0.08 indicate ‘reasonable’ approximate fit [16]. Additionally, the

expected cross-validation index (ECVI [17]) was used to compare different models for the same data, where the model with the smallest ECVI indicates the model with the best fit. For both the RMSEA and ECVI, 95 % confidence intervals were calculated using the program NIESEM [18].

Moreover, to evaluate differences between hierarchically related models, the chi-square difference test ($\text{CHISQ}_{\text{diff}}$) can be used, where a significant chi-square difference indicates a significant difference in model fit. However, to reduce the dependency on sample size and to account for model parsimony, we also considered the difference in ECVI values ($\text{ECVI}_{\text{diff}}$). The ECVI difference test can be used to test the difference in approximate model fit. The difference in ECVI is significant when the lower bound of the confidence interval around the value of the ECVI difference is larger than zero. In the specification search for measurement bias, model comparison will be evaluated using a conservative significance level of 0.1.

Results

The percentage of missing data due to attrition after the third, sixth, and 13th measurement occasions was 11, 24, and 41 %, respectively. Table 2 gives baseline and follow-up means and standard deviations of all scales used to measure HRQL, for all survival groups. Demographics and clinical characteristics for all survival groups are given in Table 3. Patients in the long-survival group show a lower pain score and higher Karnofsky score than patients in the short- and medium-survival group, indicating that these patients already have better health status at the start of treatment. There are relatively more patients with lung cancer in the short- and medium-survival groups as compared to the long-survival group, and there are relatively more women in the long-survival group as compared to the short- and medium-survival groups. This might indicate that patients with lung cancer show a more severe disease trajectory as compared to patients with breast cancer or prostate cancer. There were no significant differences between the groups with regard to age, randomization to treatment arm, number of metastases, or treatment site. Taken together, these results indicate that attrition is related to demographic and clinical characteristics that might affect HRQL. Therefore, it is important to take into account attrition when investigating changes in HRQL and possible response-shift effects.

Measurement Model

Substantive considerations and results of factor analyses were used to arrive at the *Measurement Model* in Fig. 1. The squares represent observed variables (scale scores), the circles on the top represent the common factors, functional

limitations (FUNC) and health impairments (HEALTH), and the circles at the bottom represent residual factors. Functional limitations are measured by three observed variables and health impairments are measured by six observed variables, with one observed variable in common. Classification of the common factors was based on the International Classification of Functioning, Disability and Health from the World Health Organization (WHO) that provides a framework for the description of health and health-related states [19]. In this framework, the term functioning refers to all body functions, activities, and participation, while disability refers to impairments, activity limitations, and participation restrictions. These concepts are covered by the two common factors: functional limitations (e.g., limitations of bodily functioning) and health impairments (e.g., health restrictions or symptoms). As social functioning is also considered to be an important aspect of health-related quality of life, this scale was added to the measurement and modeled to be influenced by both functional limitations and health impairments (which agrees with participation being a factor of both functioning and disability in the WHO framework).

The *Measurement Model* of Fig. 1 was the basis for a structural equation model for baseline and follow-up (T0–T12) measurements for all survival groups with no equality constraints across occasion or group (model 1.1). The χ^2 test of exact fit was significant, but the RMSEA measure indicated close fit [$\text{CHISQ}(6,193) = 11,171.87, p < .001$; $\text{RMSEA} = 0.049$, see Table 4].

No Response Shift Model

To test for the occurrence of response-shift effects, a model of no response shift was fitted to the data where all factor loadings and intercepts were constrained to be equal across measurement and survival group. The *No Response Shift Model* yielded a χ^2 test of exact fit that was significant, but the RMSEA measure indicated satisfactory fit [$\text{CHISQ}(6,466) = 12,266.14, p < .001$; $\text{RMSEA} = 0.051$]. Comparison of the *No Response Shift Model* to the *Measurement Model* shows a significant deterioration of fit [$\text{CHISQ}_{\text{diff}}(273) = 1,094.28, p < .001$; $\text{ECVI}_{\text{diff}} = 0.534$, 99.9 % CI 0.345–0.742, see Table 4], indicating the occurrence of response-shift effects.

Response Shift Model

Step-by-step modifications yielded the *Response Shift Model*, which showed several cases of response shift, as will be explained below. The fit of the *Response Shift Model* was good [$\text{CHISQ}(6,436) = 11,628.50, p < .001$; $\text{RMSEA} = 0.049$] and significantly better than the fit of the *No Response Shift Model* [$\text{CHISQ}_{\text{diff}}(30) = 637.64, p < .001$;

Table 2 Means and standard deviations of HRQL scale scores at baseline and follow-up assessments for the short-survival, medium-survival and long-survival group

	PF Mean (SD)	MB Mean (SD)	SF Mean (SD)	DP Mean (SD)	LS Mean (SD)	PA Mean (SD)	SI Mean (SD)	SY Mean (SD)
<i>Short-survival group</i>								
Baseline	3.07 (0.99)	2.13 (0.86)	2.25 (0.95)	1.98 (0.80)	2.28 (0.66)	2.77 (0.67)	1.60 (0.56)	1.46 (0.37)
Week1	3.19 (0.99)	2.26 (0.90)	2.28 (0.91)	2.01 (0.74)	2.35 (0.63)	2.65 (0.72)	1.76 (0.62)	1.49 (0.35)
Week2	3.34 (0.90)	2.49 (0.93)	2.40 (0.95)	2.03 (0.80)	2.37 (0.65)	2.54 (0.68)	1.78 (0.62)	1.47 (0.38)
<i>Medium-survival group</i>								
Baseline	2.91 (1.00)	1.99 (0.84)	2.16 (0.89)	1.96 (0.69)	2.25 (0.60)	2.73 (0.63)	1.49 (0.48)	1.44 (0.37)
Week1	3.04 (0.97)	2.08 (0.88)	2.16 (0.91)	1.87 (0.69)	2.28 (0.62)	2.52 (0.64)	1.62 (0.54)	1.47 (0.37)
Week2	3.12 (0.96)	2.15 (0.87)	2.22 (0.94)	1.86 (0.65)	2.32 (0.59)	2.51 (0.66)	1.68 (0.51)	1.46 (0.34)
Week3	3.16 (0.94)	2.23 (0.91)	2.29 (0.97)	1.84 (0.63)	2.33 (0.61)	2.42 (0.66)	1.69 (0.58)	1.46 (0.35)
Week4	3.23 (0.94)	2.30 (0.94)	2.37 (1.01)	1.91 (0.70)	2.40 (0.62)	2.47 (0.66)	1.71 (0.59)	1.45 (0.37)
Week5	3.28 (0.91)	2.40 (0.94)	2.45 (1.10)	1.96 (0.76)	2.41 (0.61)	2.43 (0.72)	1.72 (0.62)	1.51 (0.40)
<i>Long-survival group</i>								
Baseline	2.55 (1.03)	1.75 (0.79)	1.97 (0.85)	1.87 (0.68)	2.05 (0.58)	2.56 (0.65)	1.37 (0.41)	1.35 (0.31)
Week1	2.62 (1.03)	1.80 (0.83)	1.94 (0.84)	1.78 (0.64)	2.05 (0.56)	2.39 (0.61)	1.47 (0.49)	1.36 (0.31)
Week2	2.63 (1.03)	1.82 (0.85)	1.93 (0.87)	1.76 (0.66)	2.04 (0.59)	2.30 (0.62)	1.52 (0.52)	1.35 (0.29)
Week3	2.64 (1.05)	1.83 (0.84)	1.95 (0.90)	1.75 (0.67)	2.03 (0.58)	2.23 (0.60)	1.51 (0.54)	1.34 (0.30)
Week4	2.63 (1.05)	1.80 (0.83)	1.95 (0.89)	1.74 (0.69)	2.02 (0.59)	2.17 (0.64)	1.47 (0.49)	1.34 (0.30)
Week5	2.60 (1.03)	1.77 (0.83)	1.95 (0.90)	1.70 (0.66)	2.00 (0.59)	2.14 (0.63)	1.42 (0.45)	1.34 (0.30)
Week6	2.61 (1.06)	1.78 (0.83)	1.93 (0.90)	1.70 (0.67)	1.98 (0.59)	2.13 (0.64)	1.40 (0.44)	1.33 (0.30)
Week7	2.59 (1.06)	1.76 (0.84)	1.92 (0.90)	1.68 (0.67)	1.98 (0.61)	2.13 (0.63)	1.38 (0.44)	1.34 (0.32)
Week8	2.57 (1.06)	1.77 (0.86)	1.93 (0.90)	1.69 (0.69)	1.97 (0.61)	2.12 (0.66)	1.38 (0.45)	1.33 (0.32)
Week9	2.59 (1.06)	1.78 (0.86)	1.95 (0.93)	1.69 (0.67)	1.97 (0.61)	2.13 (0.65)	1.38 (0.44)	1.33 (0.33)
Week10	2.60 (1.06)	1.80 (0.87)	1.95 (0.92)	1.71 (0.71)	1.99 (0.63)	2.14 (0.67)	1.38 (0.45)	1.34 (0.34)
Week11	2.62 (1.07)	1.81 (0.87)	1.96 (0.93)	1.72 (0.72)	2.00 (0.66)	2.15 (0.68)	1.38 (0.44)	1.35 (0.34)
Week12	2.63 (1.06)	1.83 (0.91)	1.99 (0.94)	1.73 (0.73)	2.00 (0.64)	2.16 (0.68)	1.40 (0.47)	1.35 (0.34)

All scale scores range from 1 to 4. Sample size short-survival, medium-survival, and long-survival group are $n = 144$, $n = 203$, and $n = 682$, respectively

PF physical functioning, MB mobility, SF social functioning, DP depression, LS listlessness, PA pain, SI sickness, and SY treatment-related symptoms

ECVI_{diff} = 0.563, 99.9 % CI 0.412–0.732]. Although the difference in fit between the *Response Shift Model* and the *Measurement Model* is still statistically significant according to the chi-square difference test, comparison of approximate fit using the ECVI difference test indicated that the models can be considered approximately equivalent [CHISQ_{diff} (243) = 456.64, $p < .001$; ECVI_{diff} = -0.029, 99.9 % CI -0.134 to 0.100]. All parameter estimates of the *Response Shift Model* that are associated with response-shift effects are given in Tables 5 and 6.

Response shift

The step-by-step modification procedure for the detection of response shift resulted in the identification of three different response-shift effects. This procedure involved fitting a large

number of models (i.e., 24 models in the first step, 23 models in the second step, etc.) as all invariant factor loadings and intercepts were investigated iteratively in each step. Therefore, only the most relevant models are described below.

First, recalibration of ‘pain’ was detected, where the model with freely estimated intercepts of ‘pain’ for all measurements and all groups resulted in the largest improvement in model fit [CHISQ_{diff} (21) = 412.97, $p < .001$; ECVI_{diff} = 0.362, 99.9 % CI 0.244–0.501]. In addition, equality constraints on the response-shift effect estimates across groups were tenable [CHISQ_{diff} (9) = 9.50, $p = .392$; ECVI_{diff} = -0.008], indicating that recalibration of ‘pain’ is present in all survival groups to the same extent. Inspection of response-shift parameters showed that the intercept of the indicator ‘pain’ decreased over the first five measurements and stabilized around the

Table 3 Demographics and clinical characteristics of all three survival groups (N = 1,029)

	Short survival (n = 144)		Medium survival (n = 203)		Long survival (n = 682)	
	Mean (range)	SD	Mean (range)	SD	Mean (range)	SD
Age	66.13 (38–85)	9.63	64.27 (32–89)	12.09	64.23 (33–90)	11.53
Pain score*	6.52 (2–10)	2.02	6.69 (3–10)	1.95	6.05 (2–10)	2.01
Karnofsky score*	66.06 (30–100)	15.66	69.25 (20–100)	15.00	74.76 (20–100)	14.73
	N	%	N	%	N	%
<i>Type of cancer*</i>						
Breast	38	26.4	63	31.0	321	47.1
Prostate	25	17.4	33	16.3	181	27.8
Lung	53	36.8	75	36.9	106	15.5
Other	28	19.4	32	15.8	74	10.9
<i>Gender*</i>						
Male	93	64.6	119	58.6	328	48.1
Female	51	35.4	84	41.4	354	51.9
<i>Treatment arm</i>						
Single fraction	71	49.3	105	51.7	343	50.3
Multiple fractions	73	50.7	98	48.3	339	49.7
<i>Number of metastases</i>						
One	127	88.2	178	87.7	609	89.3
Two	16	11.1	21	10.3	70	10.3
Three or Four	1	0.7	3	1.5	2	0.3
<i>Treatment site</i>						
Spine	55	38.2	65	32.0	252	37.0
Pelvis	56	38.9	73	36.0	272	39.9
Femur	12	8.3	20	9.9	67	9.8
Ribs	16	11.1	19	9.4	61	8.9
Humerus	11	7.6	19	9.4	29	4.3
Other	13	9.0	33	16.3	74	10.9

* Significant differences (p < .05) between survival groups analyzed with ANOVA or chi-square test

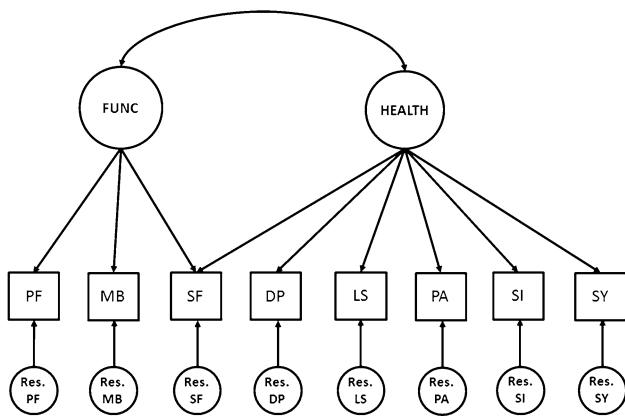


Fig. 1 Measurement Model. Circles represent latent variables (common and residual factors), and squares represent observed variables (the scale scores). *FUNC* functional limitations, *HEALTH* health impairments, *PF* physical functioning, *MB* mobility, *SF* social functioning, *DP* depression, *LS* listlessness, *PA* pain, *SI* sickness, *SY* treatment-related symptoms, *Res.* Residual. (Color figure online)

sixth measurement (see Table 5). Thus, compared to the trajectories of the other indicators of health impairments, patients in all survival groups show a stronger decrease in pain over the first 4 weeks after treatment.

Second, recalibration of ‘sickness’ was detected [CHISQ_{diff} (21) = 182.56, p < .001; ECVI_{diff} = 0.137, 99.9 % CI 0.063–0.232], where subsequent analyses showed that this response shift affected all groups to the same extent [CHISQ_{diff} (9) = 5.54, p = .785; ECVI_{diff} = -0.012]. Inspection of response-shift parameters showed that the intercept of the indicator ‘sickness’ increased over the first four measurements, after which it decreased and stabilized around the seventh measurement (see Table 5). Thus, compared to the trajectories of the other indicators of health impairments, patients in all survival groups show a temporary increase in sickness over the first 3 weeks after treatment, which diminishes again around the sixth week after treatment.

Table 4 Goodness of overall fit and difference in model fit of the models in the three-step response-shift detection procedure

Model	<i>df</i>	CHISQ	RMSEA (95 % CI)	ECVI (95 % CI)	Compared to	<i>df</i> _{diff}	CHISQ _{diff}	ECVI _{diff} (99.9 % CI)
1.0 Measurement Model	6,193	10,803.7	0.047 (0.045; 0.048)	12.32 (11.99; 12.66)	Model 1.0	273	1,462.5	0.89 (0.67; 1.14)
2.0 No Response Shift Model	6,466	12,266.1	0.051 (0.050; 0.053)	13.22 (12.86; 13.58)	Model 2.0	30	637.6	0.56 (0.41; 0.73)
3.0 Response Shift Model	6,436	11,628.5	0.049 (0.047; 0.050)	12.65 (12.31; 13.01)	Model 1.0	243	824.8	0.33 (0.17; 0.55)

N = 1,029

Table 5 Invariant parameter estimates of the Response Shift Model

HRQL scales	Intercepts (τ)	Factor loadings (<i>A</i>)	
		Functional limitations	Health impairments
Physical functioning	3.03	RS	
Mobility	2.12	0.71	
Social functioning	2.25	0.28	0.34
Depression	1.98		0.47
Listlessness	2.29		0.52
Pain	RS		0.42
Sickness	RS		0.40
Treatment-related symptoms	1.46		0.23

RS = response shift; intercepts of ‘pain’ and ‘sickness’ and factor loadings of ‘physical functioning’ are not invariant across occasions and/or groups (see Table 6); *N* = 1,029; parameter estimates are unstandardized

Third, reprioritization response shift was detected in ‘physical functioning’ with respect to functional limitations [CHISQ_{diff} (21) = 93.06, *p* < .001; ECVI_{diff} = 0.050, 99.9 % CI 0.003–0.120]. Subsequent analyses showed that this response shift was only present in the short-survival and medium-survival groups [CHISQ_{diff} (12) = 27.74, *p* = .006; ECVI_{diff} = 0.004, 99.9 % CI –0.012 to 0.046], where both groups were affected to the same degree [CHISQ_{diff} (3) = 8.16, *p* = .428; ECVI_{diff} = 0.002, 99.9 % CI –0.003 to 0.031]. Inspection of response-shift parameters shows that the factor loading of the indicator ‘physical functioning’ decreases over time (see Table 5). Thus, for patients with short or medium survival, physical functioning becomes less important for their functional limitations in the weeks after treatment, while for patients with long survival, the importance of physical functioning does not change over time.

True change

Common factor means were fixed at zero for the first measurement of the short-survival group (because of

identification requirements), so that the subsequent estimates serve as direct representations of true change compared to baseline. This also enables a meaningful comparison of common factor means across survival groups. Latent trajectories of the common factor means of all survival groups for functional and health impairments are depicted in Figs. 2 and 3, respectively. For each survival group, interrupted lines represent latent trajectories of the *No Response Shift Model* and solid lines represent latent trajectories of the *Response Shift Model*, where all response-shift effects are taken into account.

Inspection of latent trajectories of the common factor functional limitations (see Fig. 2) shows that patients with short or medium survival show significantly more limitations over time [CHISQ_{diff} (2) = 34.70, *p* < .001; CHISQ_{diff} (5) = 66.97, *p* < .001], while patients with long survival show a more constant trajectory, although there are significant changes across measurement occasions [CHISQ_{diff} (12) = 39.86, *p* < .001]. Overall, patients with short survival show more functional limitations than patients with medium or long survival, where patients with long survival show the least functional limitations. If response-shift effects would not have been taken into account, functional limitations would generally be underestimated across measurements for all survival groups (interrupted lines).

Inspection of latent trajectories of the common factor health impairments (see Fig. 3) shows that the change over time of patients with short survival is not significant [CHISQ_{diff} (2) = 3.29, *p* = .193], while patients with medium survival show significantly more impairments over time [CHISQ_{diff} (5) = 18.70, *p* = .002], and patients with long survival show significantly fewer impairments over time [CHISQ_{diff} (12) = 50.42, *p* < .001]. Again, patients with short survival show the most health impairments, and patients with long survival show the least health impairments. As can be seen from Fig. 3, if response-shift effects would not have been taken into account, health impairments would have been underestimated across measurements, but only for patients with medium or long survival (interrupted lines).

Table 6 Response-shift parameter estimates of the Response Shift Model

	T0	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
<i>Intercept 'pain'</i>													
G1	2.74	2.58	2.50										
G2	2.74	2.58	2.50	2.44	2.40	2.38							
G3	2.74	2.58	2.50	2.44	2.40	2.38	2.37	2.38	2.38	2.38	2.38	2.38	2.38
<i>Intercept 'sickness'</i>													
G1	1.54	1.66	1.71										
G2	1.54	1.66	1.71	1.71	1.68	1.65							
G3	1.54	1.66	1.71	1.71	1.68	1.65	1.63	1.63	1.63	1.63	1.61	1.60	1.61
<i>Factor loading 'physical functioning'</i>													
G1	0.87	0.80	0.71										
G2	0.87	0.80	0.71	0.68	0.64	0.62							
G3	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92

G1, G2, and G3 = short-, medium-, and long-survival groups; $N = 1,029$; parameter estimates are unstandardized

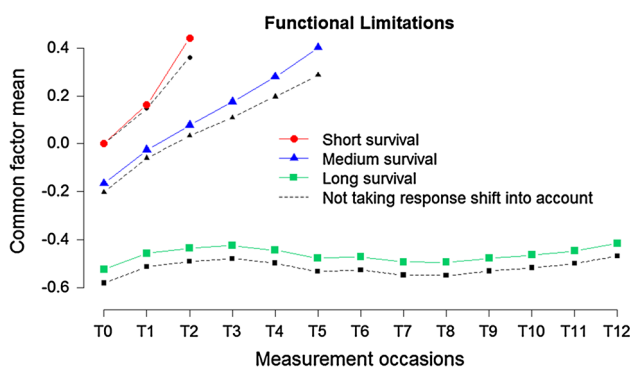


Fig. 2 Latent trajectories of functional limitations before and after accounting for response-shift effects. Red (circles), blue (triangles), and green (blocks) lines represent parameter estimates of short-, medium-, and long-survival groups, respectively; interrupted lines represent estimates of the *No Response Shift Model*, and solid lines represent estimates of the *Response Shift Model*, where all response shifts are taken into account. (Color figure online)

Discussion

We illustrated how the multigroup SEM approach can be used for the analysis of HRQL data from a longitudinal clinical trial with substantial missingness due to attrition. The approach enables the analysis of data from patients with different patterns of missing data due to attrition and thus uses more available information than standard available techniques (e.g., analysis of complete cases or complete measurements). By incorporating data from groups with different attrition rates into one analysis, this approach not only allows for detection of response shift and assessment of true change across measurements, but also allows for detection of differences in response shift and true change between groups of patients with different attrition rates.

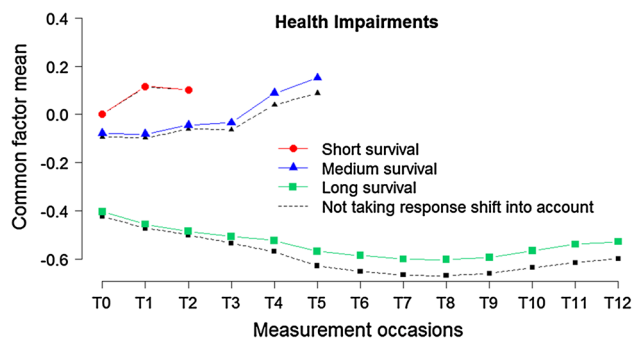


Fig. 3 Latent trajectories of health impairments before and after accounting for response-shift effects. Red (circles), blue (triangles), and green (blocks) lines represent parameter estimates of short-, medium-, and long-survival groups, respectively; interrupted lines represent estimates of the *No Response Shift Model*, and solid lines represent estimates of the *Response Shift Model*, where all response shifts are taken into account. (Color figure online)

In our sample of patients with bone metastases, we found that patients with short or medium survival show a deterioration in functional limitations, while patients with long survival show a more constant trajectory of functional limitations. For health impairments, patients with short survival show no significant change, patients with medium survival show a deterioration over time, while patients with long survival show an improvement of health impairments over time. These differential effects indicate that attrition is related to changes in HRQL and therefore emphasize the importance of taking into account missingness due to attrition when analyzing data from clinical trials. For example, when complete case analysis would have been applied to our sample of patients with bone metastases, only the long-survival group would have been considered for analysis, and the results of the short- and medium-survival groups would have been ignored. Instead, by

applying the multigroup SEM approach to distinguish survival groups, we were able to determine the impact of attrition on the results, thus enabling a more complete interpretation of findings.

In addition to changes in HRQL, we also investigated response shift. In our sample of patients with bone metastases, we detected three occurrences of response shift: recalibration response shift of the scales ‘pain’ and ‘sickness’ for all survival groups, and reprioritization response shift of the scale ‘physical functioning’ for patients with short or medium survival. Compared to the trajectories of the other indicators of health impairments, patients in all survival groups showed a stronger decrease in pain over the first 4 weeks after treatment. In addition, patients in all survival groups showed a temporary increase in sickness over the first 3 weeks after treatment, which diminished around the sixth week after treatment. Physical functioning became less important for patients’ functional limitations, but only for patients with short or medium survival. A possible explanation for the response shift in pain as a measurement of health impairments could be that the radiotherapy treatment had a larger effect on pain compared the other indicators of health impairments. In the measurement of health impairments, patients’ reporting of pain would then decrease relative to the other indicators. A possible explanation for the response shift in sickness could be that patients experienced side effects from radiotherapy and that symptoms related to sickness were relatively more prevalent than the other symptoms. As these side effects usually disappear after a few weeks, this could explain the subsequent decrease in the reporting of sickness relative to the other symptoms. A possible explanation for the response shift in physical functioning as a measurement of functional limitations could be that a declining health status coincided with a coping strategy to revalue the importance of the physical aspects of health. This could explain that this response shift only affected patients with short or medium survival. If these response shifts would not have been taken into account, functional limitations would generally be underestimated for all survival groups, whereas health impairments would generally be underestimated only for patients with medium or long survival. These occurrences of response shift and their impact on the assessment of change both emphasize the importance of investigating response shift and taking into account attrition when analyzing longitudinal data from clinical trials. A suggestion for future research would be to further investigate trends in the detected response shifts to get more insight into the development of these effects, and their impact on the changes in patients’ health-related quality of life across time.

The detection of response shift was guided by a specification search, i.e., modification of the model to improve

model fit. Such a procedure requires several decisions about model respecification. For example, a conservative significance level guards against chance findings, but might lead to overlooking interesting but statistically insignificant effects. Therefore, subjective considerations may play a role in the specification search. The current procedure focused on identifying the measurement parameter that showed the largest differences between groups, within groups over time, or both. This strategy enabled a simultaneous investigation of response shift across occasions and differences in response shift across groups. However, different strategies might be used where differences across group or over time would be prioritized over the other, and different strategies might lead to different results. For example, when different model constraints show statistically equivalent improvement in model fit, a subjective decision is required as to which constraint is indicative of response shift. This may lead to different results as choosing to release one constraint may resolve problems that also underlie the competing constraint. In our analysis, all three response shifts that were detected showed not only to be statistically significant but could also be substantively interpreted. Moreover, in each step, we considered competing models to ensure the robustness of the resulting model (i.e., that effects would not disappear in a subsequent step when another response shift was considered first). These decisions require subjective judgment, but are a necessary part of any statistical modeling procedure to ensure substantive interpretation of findings. It is therefore important to emphasize that the specification search should not only be statistically driven, but also be driven by substantive theory.

In the present study, we applied the multigroup SEM approach to take attrition into account in the investigation of changes in HRQL and detection of response-shift effects. This approach can be extended to investigate groups of patients that are distinguished based on other characteristics. For example, analyses could be based on individual (e.g., gender, age, mood, and expectations), clinical (e.g., tumor site, disease stage, and treatment), or environmental (e.g., culture and language) characteristics. Moreover, patients’ characteristics could be included in the analyses as explanatory variables to investigate their impact on HRQL trajectories. Application of statistical techniques that enable a more complete interpretation of findings will ultimately enhance our understanding of changes in HRQL, for all groups of patients.

In conclusion, when analyzing longitudinal data with substantial amounts of missing data due to attrition, it is important to make use of all available information. The multigroup SEM approach enables the analysis of data from patients with different patterns of missing data due to attrition, therefore using more available information than

standard techniques. Consequently, interpretation of findings is enhanced because possible differences between groups in the occurrence of response shift can be detected and taken into account to obtain a more valid assessment of true change. Moreover, in the assessment of true change, a comparison can be made between the latent trajectories of groups of patients with different attrition rates. Therefore, the multigroup SEM approach may be a valuable technique to advance the interpretation of findings from longitudinal clinical trials.

Acknowledgments This work was supported by a grant from the Dutch Cancer Society (Project Number: UVA 2011-4985). M. G. E. Verdam and F. J. Oort participate in the Research Priority Area Yield of the University of Amsterdam.

References

1. Sprangers, M. A. G., Moinpour, C. M., Moynihan, T., Patrick, D. L., Revicki, D. A., & The Clinical Significance Consensus Meeting Group. (2002). Assessing meaningful change in quality of life over time: A users' guide for clinicians. *Mayo Clinic Proceedings*, 77, 561–571.
2. 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, 1507–1515.
3. Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. *Quality of Life Research*, 14, 587–598.
4. Steenland, E., Leer, J., van Houwelingen, H., Post, W. J., van den Hout, W. B., Kievit, J., et al. (1999). The effect of a single fraction compared to multiple fractions on painful bone metastases: A global analysis of the Dutch Bone Metastasis Study. *Radiotherapy and Oncology*, 52, 101–109.
5. Van der Linden, Y. M., Lok, J. J., Steenland, E., Martijn, H., van Houwelingen, H., Marijnen, C. A. M., et al. (2004). Single fraction radiotherapy is efficacious: A further analysis of the Dutch bone metastasis study controlling for the influence of retreatment. *International Journal of Radiation Oncology Biology Physics*, 59, 528–537.
6. De Haes, J. C. J. M., Olschewski, M., Fayers P., Visser, M. R. M., Cull, A., Hopwood, P., & Sanderman, R. (1996). Measuring the quality of life of cancer patients with the Rotterdam Symptom Checklist (RSCL): A manual. The Netherlands: Northern Centre for Healthcare Research (NCH), University of Groningen.
7. The EuroQol Group. (1990). EuroQol: A new facility for the measurement of health-related quality of life. *Health Policy*, 16, 199–206.
8. 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, 365–376.
9. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
10. Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society (Series B)*, 39, 1–38.
11. Oort, F. J. (2001). Three-mode models for multivariate longitudinal data. *British Journal of Mathematical and Statistical Psychology*, 54, 49–78.
12. Verdam, M. G. E., Oort, F. J., van der Linden, Y. M., & Sprangers, M. A. G. (2013). *The analysis of multivariate longitudinal data: An application of the longitudinal three-mode model in health-related quality of life data*. Paper presented at the annual meeting of the Working Group Structural Equation Modeling, Bielefeld, Germany.
13. Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., et al. (2011). Openmx: An open source extended structural equation modeling framework. *Psychometrika*, 76, 306–317.
14. Steiger, J. H., & Lind, J. C. (1980). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.
15. Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173–180.
16. Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods Research*, 21, 230–258.
17. Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research*, 24, 445–455.
18. Dudgeon, P. (2003). *NIESEM: A computer program for calculating noncentral interval estimates (and power analysis) for structural equation modeling*. Melbourne: University of Melbourne, Department of Psychology.
19. World Health Organization. (2002). *Towards a common language for functioning, disability and health: The international classification of functioning, disability and health (ICF)*. Geneva: World Health Organization.