Quality of Life Research **13**: 311–320, 2004.

Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life?

A.G.E.M. de Boer[1], J.J.B. van Lanschot[2], P.F.M. Stalmeier[1,3,4], J.W. van Sandick[2], J.B.F. Hulscher[2], J.C.J.M. de Haes[1] & M.A.G. Sprangers[1]
[1]*Department of Medical Psychology;* [2]*Department of Surgery, Academic Medical Centre, Amsterdam (E-mail: a.g.deboer@amc.uva.nl);* [3]*Radiotherapy/Nijmegen Institute Cognition and Information;* [4]*The Joint Center for Radiation Oncology Arnhem-Nijmegen (RADIAN), University Center Nijmegen, Nijmegen, The Netherlands*

Accepted in revised form 28 March 2003

Abstract

Purpose: To compare the validity, reliability and responsiveness of a single, global quality of life question to multi-item scales. *Method*: Data were obtained from 83 consecutive patients with oesophageal adenocarcinoma undergoing either transhiatal or transthoracic oesophagectomy. Quality of life was measured at baseline, 5 weeks, 3 and 12 months post-operatively with a single-item Visual Analogue Scale (VAS) ranging from 0 to 100, the multi-item Medical Outcomes Study Short Form-20 (MOS SF-20) and Rotterdam Symptom Check-List (RSCL). Convergent and discriminant validity, test–retest reliability and both distribution-based and anchor-based responsiveness were evaluated. *Major findings*: At baseline and at 5 weeks, the VAS showed high correlations with the MOS SF-20 health perceptions scale ($r = 0.70$ and 0.72) and moderate to high correlations with all other subscales of the MOS SF-20 and RSCL ($r = 0.29$–0.70). The test–retest reliability intra-class correlation for the VAS was 0.87. At 5 weeks post-operatively, the distribution-based responsiveness was moderate for the VAS (standardised response mean: −0.47; effect size: −0.56), high for the physical subscales of the MOS SF-20 and RSCL (−1.08 to −1.51) and low for the psychological subscales (0.11 to −0.25). Five weeks post-operatively, anchor-based responsiveness was highest for the VAS ($r = 0.54$). *Conclusion*: The VAS is an instrument with good validity, excellent reliability, moderate distribution-based responsiveness and good anchor-based responsiveness compared to multi-item questionnaires. Its use is recommended in clinical trials to assess global quality of life.

Key words: Oesophageal neoplasms, Quality of life, Questionnaires

Introduction

Quality of life indicators have become an important outcome in clinical trials [1, 2]. There is a wide range of generic and disease-specific quality of life questionnaires covering various areas of life such as global quality of life, physical health, emotional functioning, and social life. These questionnaires differ in length varying from a single item (e.g. measuring global quality of life) to several hundred items (e.g. measuring many dimensions of quality of life) [3]. Both single-item and multi-item quality of life instruments have their strengths and weaknesses [4]. Single-item questions are the simplest approach to measure quality of life [5]. They are easier to administer [3] and less burdensome to the patients [6]. This simplicity and ease of use may result in a high rate of completed responses and operational efficiency such as data entry and data analysis. Although using a single-item global

question holds the above mentioned advantages, it may not provide all the information a researcher is interested in or a patient might feel is important to his or her quality of life. Furthermore, it might be difficult for the patient to answer a global question. Global single-item indicators require that patients consider all aspects of a phenomenon, ignore aspects that are not relevant to their situations, and differentially weigh the relevant aspects to provide a single rating [3]. Finally, clinicians might be interested in specific elements of quality of life, such as fatigue, pain or depression, which are targeted by a certain treatment. For these reasons, a single-item will often be accompanied by multi-item questionnaires.

The Visual Analogue Scale (VAS) is a type of single-item measure in which the patient indicates his or her quality of life on a line or scale, in which the anchors are usually 'best possible quality of life' and 'worst possible quality of life'. VASs have long been used in the measurement of health status and quality of life. In the early 1970s, it was applied in the scoring of the Quality of Well-Being Scale [7]. In the 1980s, Selby et al. [8] have described an instrument containing VASs for assessing quality of life in cancer patients. In the 1990s, the VAS was incorporated in some questionnaires, such as the Euroqol [9]. Recently, the VAS has been used to measure specific aspects of quality of life such as pain [10–12], incontinence [13], body image [14], and mood [15], as well as global quality of life [16, 17].

A valid questionnaire measures what it is supposed to measure and if it is reliable, it can be reproduced [1, 18]. Furthermore, instruments should be responsive. Responsiveness is the ability of an instrument to detect small but important changes. These are changes that clinicians and patients think are important or clinically relevant [18]. Multi-item questionnaires are thought to generate more reliable responses over time [3, 17] and to be more responsive to specific treatment effects [15] than single-item questions. Classical test theory has also been fairly consistent in its insistence that single-items are at a relative disadvantage to multi-item indices, because more items will allow the random error of the measurement to be cancelled out and therefore result in more reliable and precise scores [19]. However, earlier studies have shown that the global VAS is a potentially valid and reliable tool in the measurement of quality of life [9, 20, 21]. It has been linked with morbidity and mortality and can be used to screen for high-risk groups [22]. Furthermore, VASs measuring physical well-being, mood and coping [15], pain [23] and global quality of life [16, 17] have shown to be fairly responsive to clinical changes. To our knowledge, the validity, reliability and responsiveness of the VAS in a clinical trial have not yet been investigated. A valid, reliable and responsive single-item question would be of great interest for use in clinical trials, in which patients are often reluctant to fill in long questionnaires at several time points, especially when they are severely ill. In many instances they may also have concentration problems, poor eye sight and/or have difficulties writing.

The purpose of this study was to compare the validity, reliability and responsiveness of the global VAS quality of life question as compared to multi-item questionnaires. We compared the VAS with standard measures of quality of life in patients with oesophageal cancer who underwent an oesophagectomy as part of a clinical trial.

Patients and methods

Patients

Between September 1996 and September 1999, 128 consecutive patients with adenocarcinoma of the oesophagus or oesophagogastric junction were asked to participate in the present study. All patients took part in a randomised clinical trial comparing transhiatal oesophagectomy vs. transthoracic oesophagectomy with two-fields lymphadenectomy. The study was performed in the Academic Medical Center of the University of Amsterdam and University Hospital Dijkzigt of Rotterdam and approved by the medical ethics committees of both centres. Patients had to be older than 18 years and in adequate physical condition (ASA I or II). Exclusion criteria were previous or coexisting cancer, neoadjuvant chemo- or radiation therapy, and the impossibility to construct a gastric tube. All patients provided written informed consent before the operation.

Measurements

Global quality of life was measured with a VAS: a horizontal line of 100 mm ranging from 0 (worst imaginable quality of life) to 100 (perfect quality of life). Generic quality of life was measured with the MOS (Medical Outcome Studies) SF-20. The MOS SF-20 is a reliable and valid standardised measure of quality of life [24, 25]. The instrument contained 20 items measuring physical functioning, role functioning, social functioning, mental health, health perceptions and bodily pain. The MOS SF-20 was scored on a 5-point scale. All raw scales were linearly converted to a 0–100 scale, with higher scores indicating higher levels of quality of life, except for pain where higher scores indicate a higher level of pain.

Disease-specific quality of life was measured by the Rotterdam Symptom Check-List (RSCL) a self-report questionnaire designed for use with cancer patients [26]. We adapted the original RSCL by adding nine physical symptoms specific to oesophageal carcinoma and omitting seven less relevant physical items [27]. The adapted RSCL contained 40 items, covering 25 physical symptom items, seven psychological items and eight items on activities of daily living. Answers were rated on a 4-point response scale. All raw scales were linearly converted to a 0–100 scale, with higher scores indicating higher levels of quality of life.

A global rating of change in quality of life was used to classify patients according to whether they reported to have improved or deteriorated since baseline. The question was worded: 'Compared to the time just before your operation, how would you judge your quality of life at the moment?' Answers were scored on a 5-point Likert scale: (1) much better (2) somewhat better (3) no change (4) somewhat worse (5) much worse [28].

Clinical data of each patient were registered in the period after the operation until discharge. These data included (1) respiratory complications, (2) cardiac complications, (3) artificial ventilation time in days, (4) intensive care unit (ICU) stay in days, and (5) hospital stay in days.

Procedure

All patients received a postal self-report questionnaire pre-operatively (baseline), and at 5 weeks, 3, 6, 9 and 12 months after surgery. All questionnaires included the VAS, MOS SF-20, and RSCL. Post-operative assessments also included the global rating of change. Results of the measurements at baseline, 5 weeks, 3 and 12 months were chosen with regard to the validity and responsiveness data, because we expected the largest impact of the operation at 5 weeks and 3 months. At 12 months post-operatively the immediate effect of the operation is likely to have ceased.

In order to assess test–retest reliability, 48 stable patients without clinical evidence of disease recurrence were invited for an interview between 3 and 9 months after surgery. Thirty-one of these interviews took place within 3 weeks of completing a postal questionnaire (at 3, 6 or 9 months). During the interview a written self-report global VAS was assessed.

Statistics

All data were checked and analysed using the Statistical Package for the Social Sciences (SPSS-10.0). VAS scores and most (30 of 36) subscales of the MOS SF-20 and RSCL at the four measurements followed a normal distribution as judged by graphical assessment of the histogram and Q–Q plots.

Convergent and discriminant validity of the VAS, the MOS SF-20 and the RSCL were investigated by a multitrait-multimethod correlation analysis [29]. In order to compare the convergent and discriminant validity before and immediately after treatment, Pearson's correlations were calculated based on the data collected at baseline and at 5 weeks after surgery.

Correlations were considered low ($r < 0.20$), moderate ($0.20 < r < 0.50$) or high ($r > 0.50$) according to the recommendations of Cohen [30]. First, convergent validity was evaluated. The VAS should measure *overall* quality of life, consequently moderate to high correlations between the VAS and all subscales of both the MOS SF-20 and RSCL were expected. High correlations were expected between the two scales measuring overall quality of life (VAS and MOS SF-20 health perceptions), between the subscales of the MOS SF-20 and RSCL measuring physical quality of life and between the subscales of the MOS SF-20 and

RSCL measuring psychological quality of life. Second, we studied the discriminant validity. Questionnaires can be invalidated by too high correlations with other questionnaires from which they are intended to differ [29]. Therefore, we expected low correlations between physical quality of life measured by one questionnaire and psychological quality of life measured by another.

A questionnaire should yield repeatable and reproducible results when it is used repeatedly on a patient who does not experience changes in quality of life [3]. Test–retest reliability of the VAS was evaluated using the scores of disease-free patients. The intra-class correlation coefficient (ICC) by two-way analysis of variance [18] was calculated between the VAS scores of the first measurement assessed with a postal questionnaire and the second measurement assessed during the interview.

Difference scores from baseline to 5 weeks, to 3 months and to 12 months post-operatively were calculated for each quality of life measure. We examined responsiveness with two methods: with the distribution-based statistical approach and with the anchor-based approach to assess clinically relevant change [16, 31]. Two distribution-based statistics were evaluated: (1) the standardised response mean (SRM), calculated as the mean change in score divided by the standard deviation of change; and (2) the effect size (ES), equal to the mean change in score divided by the standard deviation of the baseline score [15, 18]. Both SRM and ES are calculated for each post-operative measurement, compared to baseline. SRMs and ESs were interpreted as trivial (<0.2), small (0.2–0.5), moderate (0.5–0.8) or large (>0.8) [30]. The anchor-based approach was evaluated in two ways. First, the global rating of perceived change was correlated to the difference scores of each quality of life scale at that measurement point, using Pearson's correlation coefficients. Second, the minimally important difference (MID) [28] for each questionnaire was assessed at each post-operative measurement point. The patients were divided in five groups of global change: (1) much better (2) somewhat better (3) no change (4) somewhat worse (5) much worse since baseline. A score of (2) 'somewhat better' was considered as a minimally important improvement and a score of (4) 'somewhat worse' was considered as a minimally important deterioration [28]. Differences in MIDs between the five groups of global change were evaluated with ANOVA for each quality of life measure.

Results

Sample description and baseline scores

Of the 128 patients enrolled in the study, 45 (35%) died within the first year after the operation. All remaining 83 patients (100%) completed the questionnaire at baseline, 75 patients (90%) completed the questionnaire 5 weeks after surgery, 74 patients (89%) returned the questionnaire 3 months post-operatively and 74 patients (89%) completed the questionnaire 12 months after the operation. Sixty-two patients (75%) completed all four measurements. Most patients were male (86%) and married (80%), and many had finished high-school or college (52%). The average age at baseline was 64 years (range: 44–78 years). Clinical data are reported in Table 1. Mean values of the VAS, MOS SF-20 and RSCL at the four measurements are shown in Table 2. Most scores of quality of life show a sharp decline shortly after the operation followed by a gradual improvement during the subsequent year.

Convergent and discriminant validity

The multitrait-multimethod matrix with correlations between the VAS, the MOS SF-20 and the

Table 1. Clinical characteristics of 83 patients who underwent oesophagectomy for cancer and survived more than 12 months after surgery

Characteristics	N (%)
Procedure	
THE	41 (49%)
TTE	42 (51%)
Respiratory complications	33 (40%)
Cardiac complications	16 (19%)
Artificial ventilation time in days (median (range))	1 (0–22)
ICU stay in days (median (range))	4 (1–37)
Hospital stay in days (median (range))	18 (11–64)

THE – transhiatal oesophagectomy; TTE – transthoracic oesophagectomy; ICU – intensive care unit.

Table 2. Mean values (and SD) of the VAS, MOS SF-20 and RSCL at baseline, 5 weeks, 3 and 12 months

Quality of life measure[a]	Time point			
	Baseline N = 83	5 weeks N = 75	3 months N = 74	12 months N = 74
VAS	65.4 (26)	50.6 (23)	68.9 (19)	70.3 (24)
MOS SF-20				
Physical functioning	71.8 (30)	27.4 (27)	51.2 (32)	62.6 (33)
Role functioning	75.3 (41)	14.4 (33)	52.1 (47)	71.6 (42)
Social functioning	79.5 (30)	36.7 (35)	72.4 (35)	82.7 (26)
Health perceptions	72.5 (24)	66.6 (24)	77.3 (24)	78.9 (26)
Mental health	58.1 (26)	40.9 (23)	55.7 (19)	61.4 (21)
Pain	21.7 (26)	35.3 (29)	24.3 (25)	25.3 (31)
RSCL				
Physical	85.8 (11)	71.2 (13)	78.8 (13)	82.2 (14)
Psychological	72.3 (25)	74.3 (25)	82.5 (18)	83.9 (22)
Activity	94.9 (12)	71.4 (26)	88.6 (17)	92.8 (19)

[a] Higher scores indicate better quality of life except for pain, where higher scores indicate more pain. VAS – Visual Analogue Scale; MOS SF-20 – MOS Short Form 20; RSCL – Rotterdam Symptom Check List.

RSCL is shown in Table 3. At baseline, the VAS showed moderate to high correlations with all subscales of the MOS SF-20 and RSCL ($r = 0.30$–0.67). Highest correlations with the VAS were found for the health perceptions scale of the MOS SF-20 ($r = 0.67$), physical functioning of the RSCL ($r = 0.64$) and mental health of the MOS SF-20 ($r = 0.63$). At baseline, the VAS showed the lowest correlation with pain ($r = -0.30$). Among the multi-item measures, the mental health subscale of the MOS SF-20 and the psychological functioning subscale of the RSCL correlated well with each other ($r = 0.79$).

The correlations for the scores at 5 weeks following surgery showed a similar pattern. Again, the VAS showed moderate to high correlations ($r = 0.27$–0.70) with the subscales of the MOS SF-20 and RSCL. For the VAS, highest correlation was found with health perceptions of the MOS SF-20 ($r = 0.70$). At 5 weeks post-operatively, the VAS and pain scale did correlate substantially ($r = -0.59$). Among the MOS SF-20 and RSCL, high correlations were found again for the psychological subscales ($r = 0.88$). With regard to the discriminant validity, low correlations were found between MOS SF-20 physical quality of life and RSCL psychological quality of life at baseline and at 5 weeks ($r = 0.06$ and 0.28 respectively). However, high correlations were found between MOS SF-20 mental health and RSCL physical quality of life at baseline and 5 weeks ($r = 0.54$ and 0.64, respectively).

Test–retest reliability

The mean VAS score from the postal questionnaire was 77.9 (SD 11) and the mean written VAS score at time of the interview was 78.4 (SD 13). The ICC between the VAS scores of these two measurements was 0.87 ($p < 0.01$).

Responsiveness

Difference scores since baseline, SRM and the ES of the VAS, MOS SF-20 and RSCL at 5 weeks, 3 and 12 months post-operatively are shown in Table 4. Largest differences were found after 5 weeks for physical, role and social functioning. The VAS (SRM: −0.47; ES: −0.56) and health perceptions scale of the MOS SF-20 (SRM: −0.64; ES: −0.72) showed moderate responsiveness at 5 weeks post-operatively. All physical subscales of the MOS SF-20 (physical functioning, role functioning) and RSCL (physical and activities) were highly responsive. Their SRMs varied from −0.98 to −1.24 and the ESs varied from −1.33 to −1.98. Both psychological subscales and the pain subscale showed small ESs at 5 weeks. Three months post-operatively, the SRMs and ESs of the physical subscales were still moderate, but they were small

Table 3. Multitrait-multimethod matrix for the scores of the VAS, MOS SF-20 and RSCL at baseline (upper triangle) and at 5 weeks (lower triangle)

Quality of life measure	VAS	MOS SF-20						RSCL		
		Physical functioning	Role functioning	Social functioning	Health perceptions	Mental health	Pain	Physical functioning	Psycholog. functioning	Activities
VAS	–	0.48	0.43	**0.59**	**0.67**	**0.63**	−0.30	**0.64**	0.45	**0.58**
MOS SF-20										
Physical functioning	0.38	–	**0.59**	**0.55**	0.43	0.22	−0.16	**0.58**	0.06	**0.73**
Role functioning	0.27	**0.53**	–	**0.62**	0.44	0.18	−0.16	0.48	0.06	**0.59**
Social functioning	**0.56**	**0.54**	0.46	–	**0.58**	0.31	−0.24	**0.61**	0.15	**0.68**
Health perceptions	**0.70**	0.36	0.31	**0.52**	–	0.34	−0.42	0.48	0.26	0.41
Mental health	**0.67**	0.34	0.25	0.48	**0.67**	–	−0.25	**0.54**	**0.79**	0.25
Pain	**−0.59**	−0.36	−0.13	−0.29	−0.46	−0.48	–	−0.40	−0.13	−0.01
RSCL										
Physical	**0.65**	0.36	0.25	0.45	**0.63**	**0.64**	**−0.52**	–	0.49	**0.51**
Psychological	0.47	0.28	0.17	0.37	**0.56**	**0.88**	−0.37	**0.56**	–	0.09
Activities	**0.54**	0.48	0.32	**0.53**	0.49	**0.54**	−0.24	0.39	**0.52**	–

Pearson's correlation coefficients. Correlations are considered low (<0.20), moderate (0.20–0.50) or high (>0.50: in bold).

Table 4. The difference scores since baseline, SRM and the ES of the VAS, MOS SF-20 and RSCL at 5 weeks, 3 and 12 months post-operatively

Quality of life measure	5 weeks			3 months			12 months		
	Difference since baseline	SRM	ES	Difference since baseline	SRM	ES	Difference since baseline	SRM	ES
VAS	−14.9	−0.47	−0.56[a]	3.7	0.13	0.14	2.9	0.12	0.11
MOS SF-20									
Physical functioning	−44.8	−1.24[b]	−1.51[b]	−21.8	−0.59[a]	−0.73[a]	−9.6	−0.27	−0.32
Role functioning	−58.9	−1.09[b]	−1.44[b]	−21.9	−0.42	−0.54[a]	−3.4	−0.06	−0.08
Social functioning	−43.6	−0.94[b]	−1.46[b]	−7.6	−0.22	−0.25	−7.6	0.06	0.06
Health perceptions	−17.3	−0.64[a]	−0.72[a]	−3.4	−0.13	−0.14	1.2	0.04	0.05
Mental health	−5.0	−0.21	−0.25	5.3	0.26	0.26	5.0	0.25	0.25
Pain	14.0	0.39	0.54[a]	2.7	0.09	0.10	4.7	0.14	0.18
RSCL									
Physical	−14.2	−1.08[b]	−1.33[b]	−7.2	−0.55[a]	−0.68[a]	−3.6	−0.28	−0.33
Psychological	2.8	0.13	0.11	9.2	0.46	0.37	10.0	0.55[a]	0.40
Activity	−24.0	−0.98[b]	−1.98[b]	−5.3	−0.28	−0.44	−0.4	−0.02	−0.03

SRM – standardised response mean; ES – effect size.
[a] Moderate effect; [b] Large effect.

for the other scales. One year after the operation, all responsiveness estimates were small, except for the SRM (+0.55) of the psychological functioning subscale of the RSCL.

The anchor-based approach to evaluate responsiveness examined the ability of the questionnaires to detect clinically relevant changes. Figure 1 shows the perceived global change (much better, somewhat better, no difference, somewhat worse, much worse) in quality of life per measurement point (5 weeks, 3 and 12 months). At 5 weeks post-operatively, 35 patients (47%) indicated that their quality of life had declined since baseline. However, 27 patients (36%) indicated that their quality of life had improved. Three and 12 months after the operation, many patients indicated that quality of life was better than at baseline (44%). Table 5 shows the Pearson correlation coefficients between the global ratings of perceived change in quality of life at each measurement point and the difference scores at each measurement point. Five weeks after the operation, the difference score of the VAS showed the highest correlation ($r = 0.55$) with the global rating. The difference score of the MOS SF-20 subscale role functioning had the lowest correlation. At 3 months post-surgery, the correlations between both the difference score of the RSCL-physical scale and the VAS on the one hand and the global rating of change on the other were the highest ($r = 0.32$), and at 12 months after the operation the correlation with global change was highest for the VAS ($r = 0.21$).

To establish the minimally important change, the results of the mean difference scores per global change group for the measurement at 5 weeks are shown in Table 6. In patients who were considered to have experienced a minimally important improvement (a global rating of 'somewhat better'), the overall change in VAS quality of life score was −3.3 and for a minimally important deterioration (global rating of 'somewhat worse') the change in score was −24.9. The MID for the other scales indicating improvement ranged from +12.7 to −58.3 and those indicating deterioration ranged from +3.6 to −53.6. Only the VAS, the MOS SF-20 mental health subscale and the RSCL physical subscale showed consistent results: patients in the improved global change groups have improved difference scores and the patients in the

Table 5. Pearson correlation coefficients between the global rating of perceived change in quality of life and the difference score of the VAS, MOS SF-20 and RSCL at 5 weeks, 3 and 12 months post-operatively

Quality of life measure	5 weeks	3 months	12 months
VAS	0.55**	0.32**	0.21
MOS SF-20			
Physical functioning	0.18	0.25*	0.12
Role functioning	0.15	0.30*	0.14
Social functioning	0.30**	0.15	0.10
Health perceptions	0.37**	0.28*	0.13
Mental health	0.48**	0.23	0.05
Pain	0.25*	0.15	0.02
RSCL			
Physical	0.51**	0.32*	0.15
Psychological	0.37**	0.26	0.05
Activity	0.20	0.19	0.08

*$p < 0.05$; **$p < 0.01$.

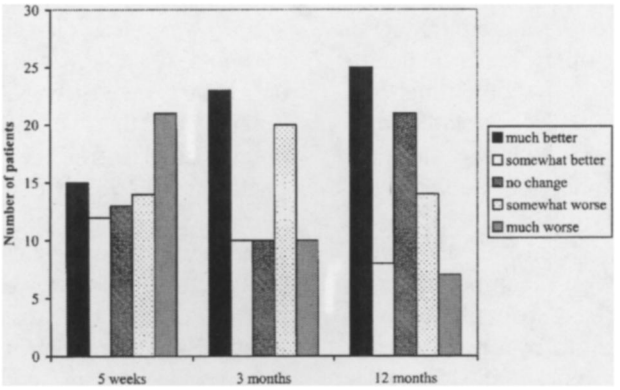

Figure 1. Global rating of perceived change in quality of life since baseline: number of patients per rating category.

Table 6. Mean difference scores since baseline of the VAS, MOS SF-20 and RSCL per global change group at 5 weeks

Quality of life measure	Global rating of change				
	Much better N = 15	Somewhat better N = 12	No change N = 13	Somewhat worse N = 14	Much worse N = 20
VAS	9.6	−3.3	−8.1	−24.9	−37.8
MOS SF-20					
Physical functioning	−46.4	−31.9	−34.6	−39.5	−60.9
Role functioning	−50.0	−58.3	−45.8	−53.6	−76.2
Social functioning	−35.7	−26.7	−21.7	−47.1	−68.6
Health perceptions	−4.8	−13.8	−5.8	−20.4	−32.9
Mental health	6.9	5.3	−0.08	−1.7	−24.4
Pain	14.3	−2.1	−4.2	17.9	31.0
RSCL					
Physical	−7.8	−6.9	−9.2	−24.4	−24.5
Psychological	11.6	12.7	2.6	3.6	−9.4
Activity	−25.8	−17.3	−9.9	−16.7	−39.9

deteriorated global change groups have worse difference scores ($p < 0.01$).

Discussion

This study investigated the validity, reliability and responsiveness of the single-item global VAS measuring quality of life as compared to multi-item questionnaires. The global VAS was found to be a valid measure of overall quality of life. As expected, the VAS showed moderate to high correlations with indicators of physical, psychological and social aspects of quality of life. Low correlations would have indicated that the global VAS did not measure quality of life at all and very high correlations with a single subscale of the MOS SF-20 and RSCL would have meant that the VAS did not assess overall quality of life but just one aspect of it. The VAS also showed the anticipated high correlation with the multi-item general health subscale of the MOS SF-20. At baseline, the overall VAS was not significantly correlated with pain, but at 5 weeks the correlation was substantial. Because the levels of pain at baseline were quite low, the overall quality of life was likely not to be influenced by it. The levels of pain were higher 5 weeks after surgery, so the impact of pain on overall quality of life has probably become larger in the period after the operation.

In an earlier study by Bernhard et al. [15] single-item indicators of physical well-being, mood and coping were correlated with standard measures at different time points. The authors found that the levels of convergent and discriminant validity of the single-item were much lower during treatment than at baseline. In the present study, this difference was not found: the VAS showed similar correlations with the standard measures at baseline and just after surgery, a finding which supports the VAS' validity as an outcome measure.

Study results differ as to whether or not single-item global questions are reliable. Some authors claim that global questions regarding quality of life can possess high reliability, while others suggest that patients may vary in the perception of their state, and thus a large random error may be associated with global questions [3, 19]. The results of this study show that a single-item global question can be reliable: test–retest reliability of the VAS proved to be very high. The test–retest reliability was based on a subgroup of clinically stable patients. One VAS was part of a postal questionnaire and one written VAS was obtained during an interview. The two different elicitation methods might have introduced error. However, despite method variation a high ICC was found. The internal consistency reliability (i.e. Cronbach's α coefficient) cannot be calculated for a single item, so no conclusions can be made on this point.

Results of the distribution-based approach showed that the physical MOS SF-20 subscales were most responsive directly after treatment at 5 weeks. The VAS showed moderate SRM and

ESs, while the mental health subscales were least responsive. The low responsiveness of the mental health subscales might suggest that the operation does not have a large emotional impact. However, the emotional scores at baseline were already low compared to normative scores of other surgery cancer patients [26], probably because patients were anxious about having cancer and having surgery. Therefore, the operation had an immediate positive effect on psychological quality of life despite the severe physical effects. Hence, a small ES was found at 5 weeks. There was a moderate ES on psychological functioning compared to baseline at 12 months, which makes this explanation plausible. The moderate ESs of the VAS might be caused by the fact that the VAS reflects both physical dimensions, which showed large SRMs and ESs, and psychological dimensions, which showed small SRMs and ESs.

The ESs found in this study were similar to other studies investigating responsiveness of VASs measuring specific areas of quality of life. Fischer et al. [23] found a SRM for a VAS measuring pain in rheumatoid arthritis patients of 0.18–1.2, depending on treatment group. In a study of breast cancer patients, Bernhard et al. investigated the responsiveness of single-item indicators 3 months after chemotherapy. The SRM for the VAS measuring physical well-being was −0.05, for the VAS mood 0.30 and for the VAS effort to cope 0.20 [15].

The present study also showed that the VAS global quality of life question was capable of measuring change from the patients' viewpoint, especially immediately after treatment. The difference score of the VAS at 5 weeks, 3 and 12 months was more strongly related to perceived change of the patients than the multi-item scales. However, we used a global rating of change, which is a retrospective method of investigating responsiveness. A number of problems might be associated with the global change question [32]. First, the reliability and validity of this question have not been assessed. Second, the judgement of change is psychologically difficult. Patients must be able to quantify both their present and initial state and then perform a mental subtraction. Therefore, this method might have biased the responsiveness results. This bias might also explain the finding that most subscales in this study showed inconsistent results with regard to the MID: patients who said that they had improved had smaller difference scores than patients who said they had not changed or had deteriorated. A limitation of the present study relates to the relatively small number of patients. As a result, the number of statistical tests performed is large as compared to the number of patients analysed.

Our findings indicate that the single-item VAS is useful. However, the choice between the single-item or multi-item measure will depend on the research question. In some studies, one would need the detail, while in others only an overall finding on a single endpoint would suffice. The choice between a single-item or multi-item scale could also depend on the underlying process to arrive at a rating. The single item relies on the patient's ability to form an overall judgement of her/his quality of life, while multi-item scales involve an external subjective or statistical weighting to combine the various elements into an overall rating. Both approaches have merit relative to the goals of the research question. Our results suggested that the VAS was indeed an overall judgement of quality of life, because the ESs of the VAS at any one time point seemed to represent a summation of the (different) ESs of the subscales at the same time point.

In conclusion, results of this study show that compared to multi-item questionnaires, the VAS is an instrument with good validity, excellent reliability, moderate estimates of distribution-based responsiveness and good anchor-based responsiveness. We recommend its use as a global quality of life measure in clinical trials.

Acknowledgements

The authors would like to thank Prof H.W. Tilanus (Department of Surgery, Erasmus University Medical Center, Rotterdam) and Prof H. Obertop (Department of Surgery of the Academic Medical Centre, Amsterdam) for their help in including the patients, and Mrs. A.F.J.W. Willems for her help in conducting part of the interviews. This project was funded by the Dutch Health Care Insurance Board, OG-96-041.

References

1. Berzon RA. Understanding and using health-related quality of life instruments within clinical research studies. In: Staquet MJ, Hays RD, Fayers PM (eds), Quality of Life Assessment in Clinical Trials. Oxford: Oxford University Press, 1998.
2. Langenhoff BS, Krabbe PFM, Wobbes T, Ruers TJM. Quality of life as an outcome measure in surgical oncology. Br J Surg 2001; 88: 643–652.
3. Fayers PM, Machin D. Quality of Life: Assessment, Analysis and Interpretation. Chichester: John Wiley and Sons, 2000.
4. Sloan JA, Aaronson NK, Cappelleri JC, Fairclough DL, Varricchio C. Assessing the clinical significance of single items relative to summated scores. Mayo Clinic Proc 2002; 77: 479–487.
5. Gill TM. Quality of life assessment: Values and pitfalls. J Roy Soc Med 1995; 88: 680–682.
6. Cunny KA, Perri M. Single-item vs. multi-item measures of quality of life. Psychol Rep 1991; 69: 127–130.
7. Patrick D, Bush J, Chen M. Methods for measuring levels of well-being for a health status index. Health Serv Res 1973; 8: 228–245.
8. Selby PJ, Chapman JA, Etazadi-Amoli J, Dalley D, Boyd NF. The development of a method for assessing the quality of life in cancer patients. Br J Cancer 1984; 50: 13–22.
9. Gudex C, Dolan P, Kind P, Williams A. Health state valuations from the general public using the Visual Analogue Scale. Qual Life Res 1995; 5: 521–531.
10. Schwenk W, Mall JW, Neudecker J, Muller JM. One visual analogue pain score is sufficient after laparoscopic cholecystectomy. Br J Surg 2002; 89: 114–115.
11. Kemler MA, Barendse GA, van Kleef M, et al. Spinal cord stimulation in patients with chronic reflex sympathetic dystrophy. N Engl J Med 2000; 343: 618–624.
12. Andren-Sandberg A, Viste A, Horn A, Hoem D, Gislason H. Pain management of pancreatic cancer. Ann Oncol 1999; 10(Suppl.): 265–268.
13. van Kampen M, de Weerdt W, van Poppel H, de Ridder D, Feys H, Baert L. Effect of pelvic-floor re-education on duration and degree of incontinence after radical prostatectomy: A randomised controlled trial. Lancet 2000; 355: 98–102.
14. O'Bichere A, Wilkinson K, Rumbles S, Norton C, Green C, Phillips RK. Functional outcome after restorative panproctocolectomy for ulcerative colitis decreases an otherwise enhanced quality of life. Br J Surg 2000; 87: 802–807.
15. Bernhard J, Sullivan M, Hürny C, Coates AS, Rudenstam C-M. Clinical relevance of single item quality of life indicators in cancer clinical trials. Br J Cancer 2001; 84: 1156–1165.
16. Bessette L, Sangha O, Kuntz KM, et al. Comparative responsiveness of generic versus disease-specific and weighted versus unweighted health status measures in carpal tunnel syndrome. Med Care 1998; 36: 491–502.
17. Sloan JA, Loprinzi CL, Kuross SA, et al. Randomized comparison of four tools measuring overall quality of life in patients with advanced cancer. J Clin Oncol 1998; 16: 3662–3673.
18. Deyo RA, Diehr P, Patrick D. Reproducibility and responsiveness of health status measures. Control Clin Trials 1991; 12: 142S–158S.
19. Gardner DG, Cumming LL, Dunham RB, Pierce JL. Single-item versus multiple-item measurement scales: An empirical comparison. Educ Psychol Meas 1998; 58: 898–915.
20. Fries JF, Ramey DR. 'Arthritis specific' global health analog scales assess 'generic' health related quality-of-life in patients with rheumatoid arthritis. J Rheumatol 1997; 24: 1697–1702.
21. Youngblut JM, Casper GR. Single-item indicators in nursing research. Res Nurs Health 1993; 16: 459–465.
22. Fayers PM, Sprangers MAG. Understanding self-rated health. Lancet 2002; 359: 187–188.
23. Fischer D, Stewart AL, Bloch DA, Lorig K, Laurent D, Holman H. Capturing the patient's view of change as a clinical outcome measure. JAMA 1999; 282: 1157–1162.
24. Stewart AL, Hays RD, Ware JE. The MOS short-form General Health Survey. Reliability and validity in a patient population. Med Care 1988; 26: 724–735.
25. Kempen GIJM. The measurement of the health status of the elderly. A Dutch version of the MOS. Tijdsch Gerontol Geriat 1992; 23: 132–140.
26. de Haes JCJM, van Knippenberg FCE, Neijt JP. Measuring psychological distress in cancer patients: Structure and application of the Rotterdam Symptom Checklist. Br J Cancer 1990; 62: 1034–1038.
27. de Boer AGEM, Onorbe Genovesi PI, Sprangers MAG, van Sandick JW, Obertop H, van Lanschot JJB. Quality of life in long-term survivors after curative transhiatal oesophagectomy for oesophageal carcinoma. Br J Surg 2000; 87: 1716–1721.
28. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. J Clin Epidemiol 1994; 47: 81–87.
29. Campbell DT, Fiske DW. Convergent and discriminant validity by the multitrait-multimethod matrix. Psychol Bull 1959; 56: 81–105.
30. Cohen J. Statistical Power Analysis for the Behavioural Sciences. New York: Academic Press, 1977.
31. Norman GR, Sridhar FG, Guyatt GH, Walter SD. Relation of distribution- and anchor-based approaches in interpretation of changes in health-related quality of life. Med Care 2001; 39: 1037–1038.
32. Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: The lesson of Cronbach. J Clin Epidemiol 1997; 50: 869–879.

Address for correspondence: A.G.E.M. de Boer, Coronel Institute for Occupational and Environmental Health, Academic Medical Center, P.O. Box 22700, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands
Phone: +31-20-566-5323; Fax: +31-20-6977161
E-mail: a.g.deboer@amc.uva.nl