Quality of Life Research **13**: 1187–1197, 2004.

Health related quality of life: A changing construct?

Jürg Bernhard[1,2], Adam Lowy[1], Natascha Mathys[1], Richard Herrmann[3] & Christoph Hürny[4,☆]
[1]*SIAK Coordinating Center, Effingerstr. 40 (E-mail: jbernhard@sakk.ch), Bern;* [2]*Institute of Med. Oncology, Inselspital, Bern;* [3]*Division of Oncology, Department of Internal Medicine, Kantonsspital, Basel;* [4]*Geriatrische Klinik, Bürgerspital, St. Gallen, Switzerland*

Accepted in revised form 24 September 2003

Abstract

In 186 patients with early colon cancer, we investigated the assumption that the meaning of 'quality of life' (QL) remains constant over time. Within a phase-III trial (SAKK 40/93), patients estimated both their overall QL and a range of disease- and treatment-related domains at five timepoints, comprising three concurrent and 2 retrospective estimates: their pre-surgery QL both before surgery and retrospectively thereafter, and their pre-adjuvant QL both at the beginning of adjuvant treatment and retrospectively about 2 months later, and their current QL 2 weeks thereafter. Multilevel models were developed to determine whether the selected domains made stable contributions to overall QL at the concurrent estimates. The weights of the domains changed over time. They did not differ significantly according to whether patients were considering their concurrent state or reflecting on this state at a later timepoint. In the process of adaptation, patients with early colon cancer substantially change the relative importance of QL domains to overall QL. This finding argues for QL as a changing construct and against the assumption that domain-specific weights are stable across distinct clinical phases.

Key words: Colon cancer, Quality of life, Reprioritization, Response shift, Then-test

Abbreviations: LASA – linear analogue self-assessment; QL – quality of life

Introduction

Probably the most fundamental feature of the quality of life (QL) construct is its subjective nature, at least in the setting of health. On this basis, patients have been given a voice in evaluating interventions. Giving the patients a voice implies a shift of reference from the health care professional to the patient. The patients' appraisal of their QL, in whatever domains, is considered valid by definition, given a valid measure.

This shift toward the patient would logically imply that the selection of relevant QL domains and their importance (i.e., weights) should also be a matter for the patient. In contrast to this desirable characteristic, almost all QL measures are designed with an externally defined set of domains. An established exception is the SEIQoL [1, 2] in which individually selected and weighted domains are aggregated into an overall score. An individual rating of the importance or the relative impact of defined domains or a group of domains has been included in various QL measures [3–8]. Its value is a matter of debate [9–14].

Whenever items are aggregated into scales or overall scores the meaning of QL is assumed to

☆ For the Swiss Group for Clinical Cancer Research (SAKK) and the Swiss Institute for Applied Cancer Research (SIAK), Bern, Switzerland.

remain constant. The use of a fixed set of domains and unchanging weights imposes this. Whether patients' weighting varies between or even within individuals to an extent that needs to be taken into account in studying QL endpoints is an open question [15]. There is increasing evidence that patients with a chronic disease may change the standards, values or conceptualization on which they base their QL estimation in the process of becoming and remaining ill [16].

We previously reported on patients with newly diagnosed colon cancer who indicated a change in internal standards in respect of QL domains [17] and health status [18] both across surgery and adjuvant treatment or observation. We described this response shift as 'reframing', signifying that patients do not assess their health against a fixed reference point (i.e., 'true' baseline) but against a frame of reference which shifts in the light of experience. Whether this process encompasses also content-related shifts is not clear.

Based on the same study, we here examined the relative importance of QL domains to overall QL at clinically different timepoints. We tested the hypothesis that the meaning of QL for patients changes across different phases of disease and treatment. In the event of a rejection of the null hypothesis of no change, we wanted to investigate whether the observed changes in domain-specific weights were attributable to a response shift, defined as change between a concurrent and the corresponding retrospective assessment.

Patients and methods

The trial

The trial (SAKK 40/93) recruited patients with radically resected and histologically proven locally advanced or nodal positive adenocarcinoma of the colon with pathologically confirmed stages (pT1-4 pN > 0 M0 and pT3-4 pN0 M0). Patients had to have a potentially curative resection (R0-resection) and no additional rectal carcinoma. The perioperative intraportal chemotherapy was a 7 day infusion of 5-fluorouracil (5-FU) starting immediately after surgery and interrupted by a 2 hour infusion of Mitomycin-C after the first 24 hours. It was to be stopped in the event of serious toxicity. If there were technical problems with the recommended intraportal catheter, switching to the intravenous route was allowed.

Randomization for post-operative adjuvant chemotherapy took place between 7 and 28 days after surgery and was recommended 2–3 days before hospital discharge. Patients were assigned to three treatment arms: observation only (A); 5-FU 450 mg/m^2 iv once weekly for 1 year plus every second week 50 mg Levamisol orally every 8 hours for 3 days (B); 5-FU 600 mg/m^2 iv once weekly for 1 year (C). Stratification variables included institution, age, tumour stage and administration of perioperative chemotherapy. Criteria for dose modifications were specified in the protocol.

The QL investigation was restricted to centres in German speaking areas of Switzerland, and to selected centres in Germany.

QL assessment

QL assessment included 14 linear analogue self assessment (LASA) indicators of components of QL. Among these, global measures of well-being and functioning were complemented by specific measures of tumour symptoms and treatment side-effects [17]. In addition, a rating for overall QL was included. All scales ranged from 0 to 100. Each patient's response on the global measures is considered a summary of the importance he or she attaches to each domain and the severity of their problems in each domain [19, 20]. This property makes these measures suitable for studying response shift or other effects of adaptation. Both specific and global LASA indicators have extensively been validated and used in cancer patients [21–28]. As shown for several global indicators, although less precise for specific symptoms or side-effects, this type of measure is responsive to the wide spectrum of reactions seen in patients on and off treatment [19].

For this analysis we prospectively selected a limited set of indicators. They had to be relevant for the different clinical situations involved (i.e., pre-/post-surgery, on/off chemotherapy) [29] and they had to reflect distinguishable concepts. In order to reduce overlap between the variables and to simplify the research question, 11 out of the 14

Table 1. Selection of QL domains

Question group	Question	Question with greatest predictive power for overall quality of life	Domain label
1	'How was your mood?' ('good' to 'bad') 'Have you suffered from fear or anxiety?' ('none' to 'great')	'How was your mood?'	Mood
2	'How was your physical well-being?' ('good' to 'bad') 'Have you been tired?' ('not at all' to 'very tired') 'How much energy did you have?' ('a lot' to 'none at all')	'How much energy did you have?'	Energy
3	'Have you been able to perform your everyday tasks?' ('very well' to 'not at all')	'Have you been able to perform your everyday tasks?'	Functional performance
4	'Have you had joint or muscle pain?' ('none' to 'great') 'Have you had pain?' ('none' to 'great')	'Have you had pain?'	Pain
5	'How good was your appetite?' ('good' to 'bad') 'How often have you felt sick or vomited?' ('not at all' to 'frequently')	'How good was your appetite?'	Appetite
6	'How much effort has it cost you to cope with your illness' ('no effort at all' to 'a great deal of effort')	'How much effort has it cost you to cope with your illness'	Coping effort

indicators were selected on the basis of previous findings [17] and clinical experience. Because of conceptual overlap, these were placed into six groups (Table 1). Within each group the indicator with the strongest independent association with overall QL was selected by multiple linear regression. All three timepoints used in the principal analysis were pooled for this purpose.

Patients were asked to estimate their pre-surgery QL both *before* surgery (surgery pre-test): '.. We would like to know how you felt during the last week before your surgery or any other treatment ...', and retrospectively *after* surgery (surgery 'then-test' [30]): '... Please think back a moment to the time before your surgery when you filled in the first questionnaire for us. Indicate on the enclosed questionnaire how you felt during the last week before your surgery ...'.

Both surgery pre- and then-test were assessed in the hospital after oral instruction by a physician or nurse, in addition to the written instructions indicated above. The surgery pre-test was to be assessed by all patients eligible for the clinical trial, regardless of whether they actually were randomized after surgery. The surgery then-test was to be completed on the day of randomization (i.e., close to hospital discharge). Sociodemographic data were also collected by the staff.

Similarly, following discharge, patients were asked to estimate their current QL *at the beginning* of randomly assigned adjuvant chemotherapy or observation (adjuvant pre-test): '... Now that you are back at home, we would like to follow-up on how you are doing ...'), and retrospectively *about 2 months later* (adjuvant then-test): '... We are interested to find out what you now think about your well-being 2 months ago ...'. Finally, patients' *current* QL under treatment or observation (adjuvant post-test) was assessed about 2 weeks after the adjuvant then-test: '... Please respond to all questions regarding how you felt during the last week ...'. The adjuvant pre-test, then-test and post-test were completed at home. Questionnaires were sent to patients with a covering letter including the relevant instructions and a stamped addressed envelope.

For all assessments, the time to be evaluated was also specified in the introductory statement to the questions on both pages of the questionnaire. For global QL the wording of the indicator was adjusted: The surgery pre-test: 'How do you rate your QL during the last week', was phrased for the surgery then-test: 'How do you rate your QL before the operation'; the adjuvant pre-test: 'How do you rate your QL during the last week' was phrased for the adjuvant then-test: 'How do you rate your QL during the time period approximately 2 months ago'.

The principal analysis was restricted to timepoints with concurrent QL assessment.

Statistical analysis

The scores of the outcome variable overall QL were skewed toward positive values and included some zeros. After square root transformation the variable was approximately normally distributed. All analyses were performed using the transformed variable.

The relative importance of the six QL domains to overall QL was analysed in multivariate multilevel models: the modelling was undertaken initially with the regression parameters for each domain constrained to be identical across the three timepoints. This assumption was then relaxed to allow for changes in the importance of each domain over time, and the resulting improvement in fit assessed using a likelihood ratio test. This analysis concerned changes between the three concurrent QL assessments, surgery pre-test, adjuvant pre-test and adjuvant post-test referred to as timepoints 1, 3 and 5.

Because some overlap remained within the six QL domains, a second analysis was limited to those domains among the six which had statistically significant independent associations with overall QL when all three concurrent assessments were considered together. Multiple linear regression models indicated that the domains mood, energy, functional performance and pain all had significant independent effects (all $p < 0.0002$). However, with these variables in the model, coping effort and appetite had non-significant effects ($p = 0.31$ and $p = 0.14$, respectively).

Clearly this alternative analysis would tend to result in the removal of domains whose importance changed substantially over time. Coping effort [24] was an example of this: coping effort had no significant independent effect on overall QL at the first timepoint but had a significant effect at the final timepoint. This might lead, in this more restricted analysis, to a false acceptance of the global null hypothesis of no change, although there is no reason for this to result in a false positive result. Because of this potential conservative bias, it was felt important to report the results of the principal analyses both of the set of six variables described above and of the subset of four.

In order to estimate the differences between the regression coefficients for each covariate at the various timepoints, multivariate two-level models (level 1 = measurement occasions, level 2 = individuals) were developed using the hierarchical modelling package MLWin [31]. For the global null hypothesis of no change between timepoints 1, 3 and 5, the model included an intercept term random at level 2 (individual) and slope terms for timepoints 3 and 5, also random at level 2, representing the changes in overall QL between timepoints 1 and 3 and timepoints 1 and 5, respectively. In addition, fixed parameters were fitted for each of the six domains. No term was random at level 1 because the two-level structure in this case is merely used to set up the multivariate structure [32]. For a detailed explanation see the corresponding methodology paper [33].

The global null hypothesis was tested by adding 12 fixed-effect (domain score $*$ timepoint) interaction terms to the model. The resulting improvement in fit was tested using a likelihood ratio test on 12 degrees of freedom. The parameter estimates for the individual interaction terms represent the estimated change in the regression coefficient for the relevant domain between the first timepoint and one or other of the later timepoints, and the p-values represent the statistical significance of the change. The analysis strategy was first to test the global null hypothesis of no change between the three concurrent QL assessments. If this was rejected, the hypotheses of no changes between timepoints 1 and 3 and between 1 and 5 were to be tested separately, domain-specific issues only being examined in cases where these hypotheses were rejected, a conventional strategy to avoid the danger of over-interpreting specific changes when there is no evidence of an overall effect.

In the event of a rejection of the null hypothesis, we wanted to investigate whether the observed changes in domain-specific weights were attributable to a response shift, defined as change between the pre-test and then-test relating to the same timepoint. We applied the same analyses as described above to the two pre-test/then-test pairs, i.e., changes between the surgery pre-test and the surgery then-test, and between the adjuvant pre-test and the adjuvant then-test. In the light of our earlier work [17, 18] we anticipated a change in domain-specific weights between the concurrent and the corresponding retrospective assessment.

We considered the possibility that patient characteristics may have differed between patients with

and without complete data, and that this may have influenced the result. A conventional approach to this is to assess whether patient characteristics differ significantly between cases with and without complete data, although this is incorrect [34]; non-significant differences do not imply comparability, nor do significant differences imply bias. Our approach was to repeat the analysis among the cases with complete data, to determine whether similar relationships between domains and QL exist in this subgroup.

Linear regression models were performed using Splus 2000 and multilevel analyses in MLWin 1.10.0006.

Results

Of the 186 patients used in this analysis, 169 (91%) had QL data available at least two of the three timepoints with concurrent assessment (1, 3 and 5). Ninety nine (53%) had QL data available at all three. The submission rates were satisfactory at each of these timepoints (75.4, 88.2 and 83.4% respectively). The characteristics of the 186 cases included in this study and of the 99 in whom QL data were complete are summarized in Table 2. Besides a minor deviation in age, the two groups were comparable.

The global null hypothesis was that the relationships between overall QL and the six domains (i.e., mood, coping effort, appetite, pain, functional performance, energy) remained constant across the three clinically distinct timepoints. Figure 1 illustrates the changes in the estimated regression parameters for each of the six domains over the three timepoints. The addition of 12 (domain score $*$ timepoint) interaction terms to the model yielded a highly significant improvement in fit (χ^2 12 df $= 35.9$, $p = 0.0003$). Thus there is strong evidence that the relationships between the domain scores and overall QL did not remain constant.

The magnitude of this effect is illustrated by the change in the proportion of the variance explained by the model when functional performance is included or excluded (186-patient model; 6 covariates). If this single domain is removed, at timepoint 1 the R^2 falls from 0.536 to 0.533, at timepoint 5 from 0.639 to 0.549. In other words, there is no change in the R^2 pre-surgery but a substantial change under adjuvant treatment or observation; the R^2 at the beginning of the adjuvant phase was 0.657.

Table 2. Patient characteristics

	Total sample N = 186 N(%)	Sample with complete data N = 99 N(%)
Sex		
Male	110 (59)	57 (58)
Female	76 (41)	42 (42)
Age		
<65	105 (56)	63 (64)
$\geqslant 65$	81 (44)	36 (36)
Type of surgery		
Right hemicolectomy/ resection of transverse colon/ sigmoid resection	124 (67)	69 (70)
Left hemicolectomy	44 (24)	22 (22)
Other	18 (10)	8 (8)
Lymph node involvement		
pN_0	97 (52)	55 (56)
$pN_1/pN_2/pN_3$	86 (46)	43 (43)
Missing	3 (2)	1 (1)
Living situation		
With spouse or partner/ with other(s)	151 (81)	84 (85)
Alone with child(ren)	2 (1)	2 (2)
Alone	32 (17)	13 (13)
Missing	1 (1)	0
Education		
No training or certificate	40 (22)	20 (20)
Training or certificate/ high school	112 (60)	59 (60)
Technical college/ academic education	30 (16)	19 (19)
Missing	4 (2)	1 (1)

For all the domains, the changes between timepoints 1 and 5 were larger than those between timepoints 1 and 3. A test of the null hypothesis of no change between timepoints 1 and 3 yielded a borderline-significant improvement in fit (χ^2 6 df $= 11.2$, $p = 0.08$) whereas the change between timepoints 1 and 5 was highly significant (χ^2 6 df $= 28.7$, $p < 0.0001$). Thus, the improvement in the fit of the model on adding the 12 interaction terms was more attributable to changes between timepoints 1 and 5 than to changes between timepoints 1 and 3.

Only the parameter estimate for energy was significantly different between the first and third

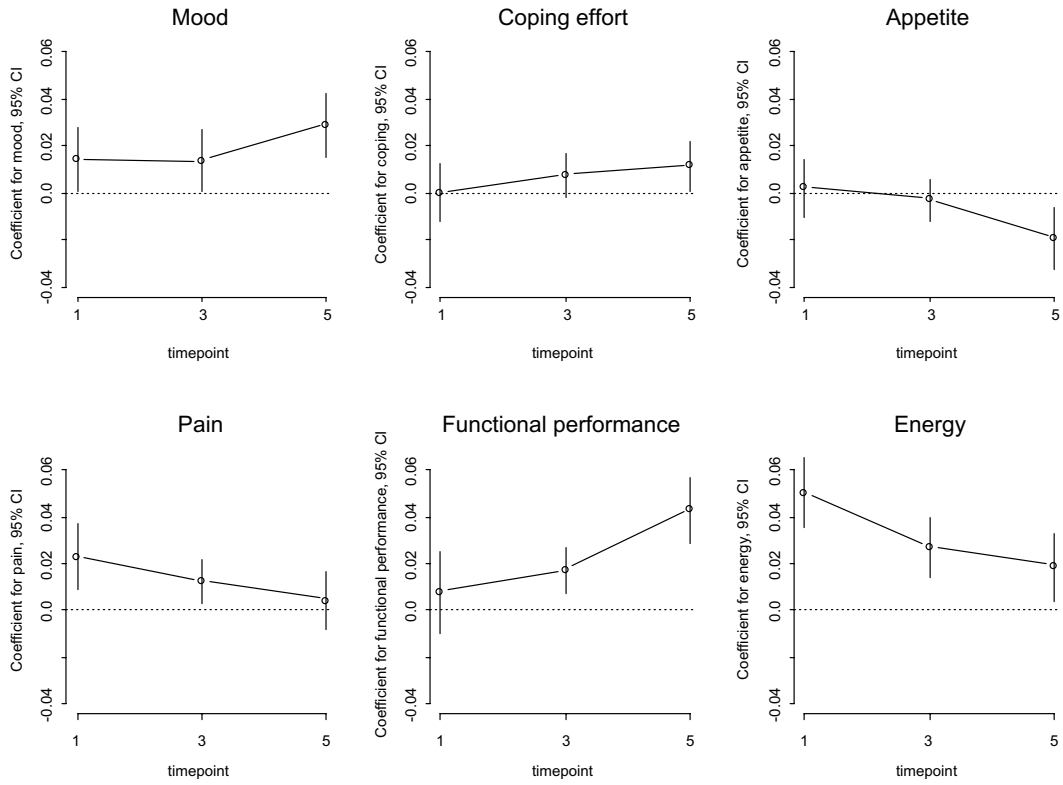

Figure 1. Changes in the estimated regression parameters on overall quality of life of the six domains over the three concurrent assessments in the total sample (n = 186).

timepoints ($p = 0.016$). The important changes between timepoints 1 and 5 were a significant increase in the parameter estimate for functional performance ($p = 0.004$) and significant decreases in those for appetite and energy ($p = 0.02$ and $p = 0.003$, respectively).

We explored the impact of adjuvant therapy on these changes by comparing patients assigned to chemotherapy with those assigned to observation only. The pattern was not consistent (data not shown) and not further investigated due to the limited power for this potential higher order effect.

A similar pattern of changes was found when the model was restricted to the four domains which had significant independent associations with overall QL (mood, pain, functional performance and energy). With this restricted model, the improvement in fit on adding the eight relevant interaction terms was again highly significant ($p < 0.001$). Again there was a significant reduction in the regression coefficient for energy at timepoint 3 compared with 1 ($p = 0.012$). The changes between timepoints 1 and 5 were more striking, including a significant increase in the parameter estimate for functional performance ($p = 0.008$), and significant falls in the parameter estimates for pain ($p = 0.046$) and energy ($p < 0.001$) between timepoints 1 and 5.

The analyses thus far estimated group average domain weights, whereas it is likely that domain weights vary from individual to individual, as well as over time. We explored this by fitting domain scores as random effects. The resulting improvements in fit for individual domains were either of borderline or no significance. Because of the relatively small size of the dataset and the small number of timepoints this should not be interpreted as an indication that individual variation in domain weights does not exist, but in the interests of simplicity, the results of the more parsimonious models with domains as fixed effects are presented.

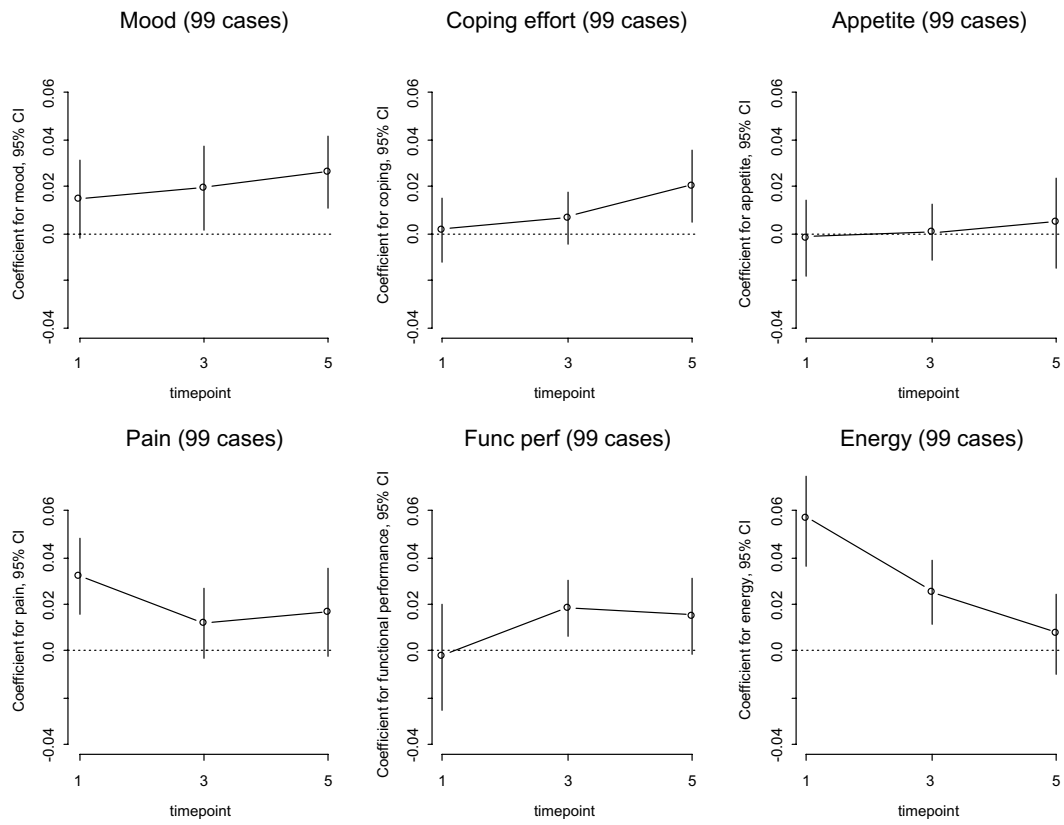

Figure 2. Changes in the estimated regression parameters on overall quality of life of the six domains over the three concurrent assessments in the sample with complete data (n = 99).

Because QL data were frequently missing, it is possible that part of the change in the conception of QL between the timepoints is due to a different set of patients being included at each timepoint. To check for this effect, the principal analysis was repeated within the 99 cases for whom data were complete. The changes in the parameter estimates of the six variables were broadly similar to the changes observed in the full sample of 186 cases, as shown in Figure 2, although the confidence intervals were considerably wider due to the substantially reduced sample size. The test of the global null hypothesis was non-significant ($\chi^2 12$ df = 17.6, $p = 0.129$), although this could be due to the reduced sample size and does not contradict the main result.

Finally, we investigated whether the observed changes in domain-specific weights were attributable to a response shift, defined as change between the pre-test and then-test. We analysed the surgery pre-test and then-test, and the adjuvant pre-test and then-test in an analogous way to the previous analyses. There was no evidence to suggest that the parameters for the six domains changed between the (concurrent) surgery pre-test and the (retrospective) then-test ($p = 0.24$), nor between the adjuvant pre-test and the corresponding then-test 2 months thereafter ($p = 0.51$). The changes were also non-significant when the models were restricted to the four domains mood, pain, functional performance and energy (surgery: $p = 0.11$; adjuvant phase: $p = 0.26$).

Discussion

In this study we investigated the assumption that the meaning of 'QL' remains constant for patients with early colon cancer, by analysing the relative

importance of QL domains to overall QL across distinct clinical phases.

The principal result is a rejection of the null hypothesis that this group of patients was characterized by an unchanging set of relationships between mood, coping effort, appetite, pain, functional performance and energy and their overall QL: patients substantially changed the relative importance of these domains across surgery and the adjuvant phase. This suggests that the meaning of QL for these patients changed.

Importantly, the changes in individual weights of domains had a strong tendency to be monotonic, the changes over the first 2 months of the adjuvant phase almost all being in the same direction as those between pre-surgery and beginning of the adjuvant phase. This pattern suggests a continuous process of adaptation over surgery, hospital discharge, rehabilitation and follow up with or without chemotherapy. The clinical course is characterized by an increasing controllability of symptoms and side-effects. Similarly, expectations rise, for example of functional performance, which are associated with QL estimates [35, 36].

The explanation for the change in direction of energy is more elusive. In the surgery recovery phase patients are less bothered by disease-related concerns and feel more energetic with increasing time. They are faced with finding their way back into daily life [29]. This shift may explain the increasing and dominant weight of functional performance and the decreasing weight of energy and perhaps pain in the adjuvant phase.

Our anticipation that these changes were attributable to a response shift defined as change between the concurrent (pre-test) and retrospective (then-test) assessment was not supported. For both surgery and the adjuvant phase, there was no evidence that the relative weights of the six domains differed between the two corresponding estimates. Patients' perception of the past situation in terms of weighting appeared to be not contaminated by their present status of adaptation.

Given the number of hypotheses tested in these analyses, the *p*-values for changes in individual domains must be interpreted with caution. Nevertheless some changes were considerably greater than would be expected by chance, namely an increasing importance of functional performance in the patients' judgment of their QL, and an equally strong tendency for the importance of energy to decrease. These changes were found whether or not coping effort and appetite were included in the models, and in both cases the changes over the three concurrent assessments suggested a continuous process.

Large changes in the parameter estimates and in the *p*-values for individual domains occur frequently by chance. Informal comparisons are apt to yield misleading results. The statistical approach used in this analysis avoids the need to make subjective judgments about the importance of changes in parameter estimates and *p*-values. It allowed the global null hypothesis of no change in the domain-specific regression coefficients over time to be formally tested, and also allowed for testing of the statistical significance of changes in the regression coefficients of individual domains. The parameter estimates, confidence intervals and *p*-values for the interaction terms provide estimates of the magnitude and direction of changes in the importance of individual domains in the patients' perception of QL, the extent of uncertainty around this estimate and the probability of observing a change as great or greater than this by chance, respectively.

We faced a considerable proportion of missing data. The process of missingness may be non-ignorable [37]. The chosen group-level comparisons over time reflect differences in the patient populations included at the different timepoints as well as true within-individual changes. It is reassuring that a similar pattern was seen when the analysis was restricted to patients with complete data, because changes seen in this analysis can only be attributable to a genuine change of weights. However, this group of patients may not be representative, and these analyses involved a substantial reduction in sample size and statistical power. In principle our result may have been influenced by the change in setting of questionnaire administration between timepoint 1 (in hospital) and timepoints 3 and 5 (at home). However, important changes occurred between the latter two timepoints in which the questionnaires were applied in the same way, so it is unlikely that the result is an artefact of this type. Again, in principle the changes could be influenced by slight changes in the wording of the questionnaire, but since the

questionnaire items at timepoints 3 and 5 were identical this is unlikely to explain the whole result.

Our findings argue for QL as a changing construct and against the assumption that domain-specific weights are stable across clinically distinct phases. However, these findings are confined to the selected domains and response format. We did not use a direct, explicit weighting by the patients. We do not see this result as invalidating global ratings, or even changes in global ratings over time. However, if confirmed, it would weaken the validity of global scores which are constructed from domains using fixed weights. The impact of changing weights on the operationalization of QL endpoints should be further investigated in longitudinal designs. In particular, the affective component of QL is associated with patient's physical situation and prone to change [38].

The value of direct individual weighting is a different although related issue. It has been discussed since the beginning of QL research [3–14] and involves different concepts, with measures of 'objective' health status and 'subjective' QL as the extremes of a continuum. The assumption of stable weights has been the basis for the development of most QL measures and for normative comparisons across clinically different populations. In contrast, individual preferences are part of the paradigm of utility measures used in decision models. Assigning individually assessed preference weights to self-reported level of functioning did not result in stronger relationships with utilities [39]. In another study, if heterogeneity across patients was accounted for, considerably more variation of time trade-off values could be explained by QL domains [40]. As in our study, this finding was based not on weights elicited from the patients but on weights inferred by statistical analyses.

The question of changing weights is also of clinical interest. Assessing specific symptoms and side-effects is complementary to more global measures of function and well-being [41]. This concept has been adopted by various QL questionnaires. Studying the associations among the different measures in relation to the biomedical variables can give further insight and is helpful in defining risk factors for poor adjustment and in developing intervention strategies [42]. To what extent such associations change across clinically different situations is a question to be investigated.

In summary, patients with colon cancer substantially changed the relative importance of domains to their overall QL across surgery and adjuvant treatment. This finding argues for QL as a changing construct and against the assumption that domain-specific weights are stable across distinct clinical phases. Its methodological and clinical relevance needs to be investigated with different QL measures in different populations and situations of clinical transitions.

Acknowledgements

This study needed considerable persuasiveness and logistic efforts. We would like to thank the patients who took the time to complete the questionnaires, the data managers, nurses and surgeons who took care of this study in the centres, Heidi Gusset for central data management, and Brigitte van Wegberg for trial coordination at the beginning.

References

1. McGee HM, O'Boyle CA, Hickey A, O'Malley K, Joyce CRB. Assessing the quality of life of the individual: The SEIQoL with a healthy and a gastroenterology unit population. Psychol Med 1991; 21: 749–759.
2. Hickey AM, Bury G, O'Boyle CA, Bradley F, O'Kelly FD, Shannon W. A new short form individual quality of life measure (SEIQoL-DW): Application in a cohort of individuals with HIV/AIDS. Brit Med J 1996; 313: 29–33.
3. Ferrans CE, Powers MJ. Quality of life index: Development and psychometric properties. ANS Adv Nurs Sci 1985; 8: 15–24.
4. Tugwell P, Bombardier C, Buchanan WW, Goldsmith CH, Grace E, Hanna B. The MACTAR Patient Preference Disability Questionnaire – an individualized functional priority approach for assessing improvement in physical disability in clinical trials in rheumatoid arthritis. J Rheumatol 1987; 14: 446–451.
5. Cella DF, Tulsky DS, Gray G, et al. The functional assessment of cancer therapy scale: Development and validation of the general measure. J Clin Oncol 1993; 11: 570–579.
6. Becker M, Diamond R, Sainfort F. A new patient focused index for measuring quality of life in persons with severe and persistent mental illness. Qual Life Res 1993; 2: 239–251.
7. Henrich G, Herschbach P. Questions on Life Satisfaction (FLZM) – A short questionnaire for assessing subjective quality of life. Eur J Psychol Assessment 2000; 16: 150–159.
8. Wright JG, Young NL. The patient-specific index: Asking patients what they want. J Bone Joint Surg Am 1997; 79: 974–983.

9. Mozes B, Shabtai E. The contribution of personal rating to the clinimetric functioning of a generic quality of life instrument. J Clin Epidemiol 1996; 49: 1419–1422.
10. Larsson G, Sjoden PO, Oberg K, von Essen L. Importance-satisfaction discrepancies are associated with health-related quality of life in five-year survivors of endocrine gastrointestinal tumours. Ann Oncol 1999; 10: 1321–1327.
11. Tishelman C, Degner LF, Mueller B. Measuring symptom distress in patients with lung cancer. A pilot study of experienced intensity and importance of symptoms. Cancer Nurs 2000; 23: 82–90.
12. Trauer T, Mackinnon A. Why are we weighting? Qual Life Res 2001; 10: 579–585.
13. Osman LM, McKenzie L, Cairns J, et al. Patient weighting of importance of asthma symptoms. Thorax 2001; 56: 138–142.
14. Neudert C, Wasner M, Borasio GD. Patients' assessment of quality of life instruments: A randomised study of SIP, SF-36 and SEIQoL-DW in patients with amyotrophic lateral sclerosis. J Neurol Sci 2001; 191: 103–109.
15. Sprangers MA, Schwartz CE. The challenge of response shift for quality-of-life-based clinical oncology research (editorial). Ann Oncol 1999; 10: 747–749.
16. Sprangers MAG, Schwartz CE. Integrating response shift into health-related quality of life research: A theoretical model. Soc Sci Med 1999; 48: 1507–1515.
17. Bernhard J, Hurny C, Maibach R, Herrmann R, Laffer U. Quality of life as subjective experience: Reframing of perception in patients with colon cancer undergoing radical resection with or without adjuvant chemotherapy. Swiss Group for Clinical Cancer Research (SAKK). Ann Oncol 1999; 10: 775–782.
18. Bernhard J, Lowy A, Maibach R, Huerny C. Response shift in the perception of health for utility evaluation. An explorative investigation. Eur J Cancer 2001; 37: 1729–1735.
19. Bernhard J, Sullivan M, Huerny C, Coates AS, Rudenstam CM. Clinical relevance of single item quality of life indicators in cancer clinical trials. Br J Cancer 2001; 84: 1156–1165.
20. Bernhard J, Maibach R, Thuerlimann B, Sessa C, Aapro MS. Patients' estimation of overall treatment burden: Why not ask the obvious? J Clin Oncol 2002; 20: 65–72.
21. Coates A, Fisher Dillenbeck CF, MCNeil DR, et al. On the receiving end II. Linear analogue self-assessment (LASA) in evaluation of aspects of the quality of life of cancer patients receiving chemotherapy. Eur J Cancer Clin Oncol 1983; 19: 1633–1637.
22. Coates A, Gebski V, Bishop JF, et al. Improving the quality of life during chemotherapy for advanced breast cancer. A comparison of intermittent and continuous treatment strategies. N Engl J Med 1987; 317: 1490–1495.
23. Butow P, Coates A, Dunn S, Bernhard J, Hürny C. On the receiving end IV: Validation of quality of life indicators. Ann Oncol 1991; 2: 597–603.
24. Hürny C, Bernhard J, Bacchi M, et al. The Perceived Adjustment to Chronic Illness Scale (PACIS): A global indicator of coping for operable breast cancer patients in clinical trials. Swiss Group for Clinical Cancer Research (SAKK) and the International Breast Cancer Study Group (IBCSG). Support Care Cancer 1993; 1: 200–208.
25. Hürny C, Bernhard J, Coates AS, et al. Impact of adjuvant therapy on quality of life in women with node-positive operable breast cancer. International Breast Cancer Study Group. Lancet 1996; 347: 1279–1284.
26. Hürny C, Bernhard J, Coates A, et al. Responsiveness of a single-item indicator versus a multi-item scale: Assessment of emotional well-being in an international adjuvant breast cancer trial. Med Care 1996; 34: 234–248.
27. 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.
28. Bernhard J, Thurlimann B, Schmitz SF, et al. Defining clinical benefit in postmenopausal patients with breast cancer under second-line endocrine treatment: Does quality of life matter? J Clin Oncol 1999; 17: 1672–1679.
29. Bernhard J, Hürny C. Gastrointestinal cancer. In: Holland J (ed.), Psycho-oncology, Oxford: Oxford University Press, 1998; 324–339.
30. Howard GS, Ralph KM, Gulanick NA, Maxwell SE, Nance SW, Gerber SK. Internal invalidity in pretest-posttest self-report evaluations and a re-evaluation of retrospective pretests. Appl Psychol Meas 1979; 3: 1–23.
31. Rasbash J, Browne W, Goldstein H. A user's guide to MLWIN. London: Centre for Multilevel Modelling, Institute of Education, University of London, 2000.
32. Goldstein H. Multilevel statistical models. London: Edward Arnold, 1995.
33. Lowy A, Bernhard J. Quantitative assessment of changes in patients' constructs of quality of life: an application of multilevel models. Qual Life Res 2004; 13: 1177–1185.
34. Rothman K. The assessment and control of confounding. Modern epidemiology. Boston: Little, Brown and Co, 1986; 125–126.
35. Wan GJ, Counte MA, Cella DF. The influence of personal expectations on cancer patients' reports of health-related quality of life. Psychooncology 1997; 6: 1–11.
36. Bromberg MB, Forshew DA. Comparison of instruments addressing quality of life in patients with ALS and their caregivers. Neurology 2002; 58: 320–322.
37. Little R, Rubin D. Statistical analysis with missing data. New York: Wiley, 1987.
38. de Haes JC, de Ruiter JH, Tempelaar R, Pennink BJ. The distinction between affect and cognition in the quality of life of cancer patients – sensitivity and stability. Qual Life Res 1992; 1: 315–322.
39. Jansen SJ, Stiggelbout AM, Nooij MA, Kievit J. The effect of individually assessed preference weights on the relationship between holistic utilities and nonpreference-based assessment. Qual Life Res 2000; 9: 541–557.
40. Bult JR, Hunink MG, Tsevat J, Weinstein MC. Heterogeneity in the relationship between the time tradeoff and Short Form-36 for HIV-infected and primary care patients. Med Care 1998; 36: 523–532.
41. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993; 85: 365–376.

42. Pehrsson K, Olofson J, Larsson S, Sullivan M. Quality of life of patients treated by home mechanical ventilation due to restrictive ventilatory disorders. Respir Med 1994; 88: 21–26.

Address for correspondence: Jürg Bernhard, SIAK Coordinating Center, Effingerstr. 40, 3008 Bern, Switzerland
Phone: +41-0-31-389-91-91; Fax: +41-0-31-389-92-00
E-mail: jbernhard@sakk.ch