British Journal of Anaesthesia **84 (1)**: 11–15 (2000)

Validity and reliability of a postoperative quality of recovery score: the QoR-40†

P. S. Myles*, B. Weitkamp, K. Jones, J. Melick and S. Hensen

Department of Anaesthesia and Pain Management, Alfred Hospital, Commercial Road, Prahran, Victoria 3181, Australia

**Corresponding author*

Quality of recovery after anaesthesia is an important measure of the early postoperative health status of patients. We attempted to develop a valid, reliable and responsive measure of quality of recovery after anaesthesia and surgery. We studied 160 patients and asked them to rate postoperative recovery using three methods: a 100-mm visual analogue scale (VAS), a nine-item questionnaire and a 50-item questionnaire; the questionnaires were repeated later on the same day. From these results, we developed a 40-item questionnaire as a measure of quality of recovery (QoR-40; maximum score 200). We found good convergent validity between QoR-40 and VAS ($r=0.68$, $P<0.001$). Construct validity was supported by a negative correlation with duration of hospital stay (rho=–0.24, $P<0.001$) and a lower mean QoR-40 score in women (162 (SD 26)) compared with men (173 (17)) ($P=0.002$). There was also good test–retest reliability (intra-class $r_i=0.92$, $P<0.001$), internal consistency (Cronbach's $\alpha=0.93$, $P<0.001$) and split-half coefficient ($\alpha=0.83$, $P<0.001$). The standardized response mean, a measure of responsiveness, was 0.65. The QoR-40 was completed in less than 6.3 (4.9) min. We believe that the QoR-40 is a good objective measure of quality of recovery after anaesthesia and surgery. It would be a useful end-point in perioperative clinical studies.

Br J Anaesth 2000; **84**: 11–15

Keywords: anaesthesia, audit; measurement techniques, recovery; recovery, postoperative; complications, morbidity; research, anaesthesia; surgery, postoperative period

Accepted for publication: June 30, 1999

Measurement of patient health status, or quality of life, has become an important end-point in clinical studies as it represents, in part, the patient's perception of their outcome of care.[1–4] This approach has rarely been used in anaesthesia and surgery studies[5] which have focused traditionally on other recovery indices, such as time to awakening, duration of stay or various adverse sequelae such as pain, emesis or confusion.[5–10] Although these events are usually transient, they are of major concern to many patients and often leave them with negative recollections of their recovery from surgery.[11–13] Poor quality recovery frequently prolongs duration of stay in the recovery room or delays discharge from hospital, both of which have significant implications for resource utilization.

We have previously developed a nine-item questionnaire to measure quality of recovery after general anaesthesia and surgery.[14] Psychometric evaluation revealed moderate validity and reliability (coefficients 0.50–0.61), suggesting it was an acceptable instrument for *group* measurements. The aim of this study was to develop an expanded questionnaire and test its validity, reliability and responsiveness in hospital surgical practice.

Patients and methods

After obtaining approval from the Ethics Committee and informed consent, we studied male and female patients, aged more than 18 yr, undergoing general anaesthesia and surgery. Patients were excluded if they had poor English comprehension, psychiatric disturbance that precluded complete cooperation, known history of alcohol or drug dependence, or any severe pre-existing medical condition that limited objective assessment after operation.

Baseline (preoperative) data were collected and patients were asked to complete two questionnaires. The first had nine items asking the patient to rate their status on a three point scale, the QoR score.[14] The second was a more comprehensive 50-item questionnaire with items on a five-point Likert scale (for positive items, 1='none of the time' to 5='all of the time'; for negative items the scoring was reversed). (The full questionnaire is available on request from the corresponding author.) The 50 items chosen had been identified previously by patients, their relatives, nursing

†The article is accompanied by Editorial I.

and medical staff as important during the postoperative recovery period.[14] We also collected other perioperative data, such as details of surgery and anaesthesia, duration of stay in the recovery room and total hospital stay.

On the morning after surgery, patients were asked to rate their overall postoperative recovery using a 100-mm visual analogue scale (VAS), marked from 'poor recovery' to 'excellent recovery', as an alternative overall assessment of recovery,[1 14] and then complete the QoR score and 50-item questionnaire. Patients were instructed to record the time taken to complete the 50-item questionnaire and to repeat the QoR score and 50-item questionnaire several hours later (as a measure of repeatability). Inpatients were instructed to place the second set of completed questionnaires in an internal mail envelope. Patients who were discharged home on the day of surgery were given the questionnaires and instructed to complete them the following day, and then return the completed questionnaires in a self-addressed envelope provided.

The 50-item questionnaire was later analysed to remove items that were not correlated with quality of recovery (using the total score), as identified by a Pearson correlation coefficient <0.30. Items removed were: ability to read, hiccups, pain at the needle or i.v. site, cough, leg cramps, dry mouth, constipation, diarrhoea and difficulty with micturition ('trouble with urine'). This resulted in a 40-item questionnaire intended to measure quality of recovery (the QoR-40). The items were then grouped according to various aspects (dimensions) of recovery: emotional state ($n=9$), physical comfort ($n=12$), psychological support ($n=7$), physical independence ($n=5$) and pain ($n=7$) (Table 1).

Table 1 Dimensions of the QoR-40 identified to represent aspects of good quality recovery after anaesthesia and surgery. Positive items were scored from 1 (worst) to 5 (best); scores were reversed for negative items. Internal consistency (Cronbach's α) and item-to-own dimension correlation coefficients are presented

	Coefficient
Emotional state ($\alpha=0.82$)	
Q.2 Feeling comfortable	0.72
Q.9 Having a general feeling of well-being	0.74
Q.12 Feeling in control	0.67
Q.28 Bad dreams	0.45
Q.36 Feeling anxious	0.74
Q.37 Feeling angry	0.63
Q.38 Feeling depressed	0.71
Q.39 Feeling alone	0.66
Q.40 Difficulty falling asleep	0.61
Physical comfort ($\alpha=0.83$)	
Q.1 Able to breathe easy	0.60
Q.5 Have a good sleep	0.67
Q.10 Being able to enjoy food	0.61
Q.11 Feeling rested	0.59
Q.19 Nausea	0.71
Q.20 Vomiting	0.52
Q.21 Dry retching	0.63
Q.24 Feeling restless	0.65
Q.25 Shaking or twitching	0.48
Q.26 Shivering	0.60
Q.27 Feeling too cold	0.51
Q.34 Feeling dizzy	0.60
Psychological support ($\alpha=0.83$)	
Q.13 Able to communicate with hospital staff (when in hospital)	0.64
Q.14 Able to communicate with family or friends	0.78
Q.15 Getting support from hospital doctors (when in hospital)	0.67
Q.16 Getting support from hospital nurses (when in hospital)	0.72
Q.17 Having support from family or friends	0.66
Q.18 Able to understand instructions or advice	0.67
Q.35 Feeling confused	0.59
Physical independence ($\alpha=0.80$)	
Q.3 Able to return to work, or usual home activities	0.67
Q.4 Able to write	0.74
Q.6 Have normal speech	0.52
Q.7 Able to wash, brush teeth or shave	0.85
Q.8 Able to look after own appearance	0.88
Pain ($\alpha=0.77$)	
Q.22 Moderate pain	0.59
Q.23 Severe pain	0.60
Q.29 Headache	0.60
Q.30 Muscle pains	0.79
Q.31 Backache	0.70
Q.32 Sore throat	0.63
Q.33 Sore mouth	0.69

Validity testing

Validity is a measure of accuracy and was difficult to verify because there were no accepted alternative methods of measuring quality of postoperative recovery (we chose not to use the recently developed QoR score[14] because it had common items and so would have a spuriously high correlation). We used the following to evaluate validity:

(1) Convergent validity: we compared the QoR-40 with the VAS, and also measured inter-item correlations.

(2) Construct validity: we compared the QoR-40 score between men and women, as women were expected to have a poorer quality of recovery on the basis of previous studies.[6–9] We also measured the associations between the QoR-40 and the time required for completion of the questionnaire, duration of stay in the recovery room and duration of hospital stay.

Reliability

Reliability is a measure of consistency and was assessed by:

(1) Test–retest reliability: we asked patients to complete both questionnaires on a second occasion, later on the same postoperative day.

(2) Internal consistency of the QoR-40: we also measured the median correlation between items within each dimension and the item-to-own dimension correlations.

(3) Split-half reliability: we measured the correlation between split segments of the QoR-40.

Clinical acceptability and responsiveness

A health status instrument should be acceptable to patients and staff and be able to detect a meaningful change in health status.[15–17] We used the following:

(1) Recruitment rate.

(2) Time taken for patients to complete the 50-item questionnaire.

(3) Successful completion and return rate.

(4) Responsiveness: we measured standardized response means.[16]

Table 2 Patient characteristics, extent of surgery and timing of assessment (n=160) (mean (SD or range), median [IQR] or number (%))

Age (yr)	44 (18–81)
Sex (M/F) (% male)	75/85 (47)
ASA status (n (%))	
I	71 (44)
II	72 (45)
III	17 (11)
Extent of surgery (n (%))	
Day-case	25 (16)
Minor	78 (49)
Major	57 (36)
Type of surgery (n (%))	
General	48 (30)
Gynaecology	33 (21)
Orthopaedic	25 (16)
Ear, nose and throat	22 (14)
Urology	15 (9)
Other	17 (10)
Duration of surgery (min)	70 [45–120]
Recovery room stay (min)	60 [45–84]

Statistical analysis

Data are presented as mean (SD), median (interquartile range (IQR)) and 95% confidence intervals (95% CI). Associations were measured using Pearson correlation coefficients (r), Spearman rank correlation (rho) or Cronbach's alpha (α); test–retest reliability (concordance) was measured using the intra-class correlation coefficient (r_i).[15] Repeatability was also calculated from the within-subjects SD, based on the Bland–Altman method.[18] Comparisons between correlation coefficients were based on the normal approximation. Changes from baseline were compared using the paired t test. Standardized response means were calculated as the mean change in score divided by its SD.[16] All analyses were performed using SPSS for Windows v8.0 (SPSS Inc., Chicago, IL, USA). The null hypothesis was rejected if two-tailed $P<0.01$.

Results

Of the 192 patients approached in this study, there was only one refusal (recruitment rate >99%); another 26 patients who had been discharged from hospital did not return their questionnaires (completion and return rate 87%). Three patients were excluded; two because their surgical procedure was performed under local anaesthesia and one because of heavy sedation in the intensive care unit. Therefore, there were 160 evaluable patients (age range 18–81 yr) recovering from most types of surgery (Table 2). Neuromuscular blocking agents were used on 111 (69%) occasions. Median hospital duration of stay was 3 (IQR 2–6) days. The questionnaires were completed in hospital by 118 (74%) patients and at home by 42 (26%).

The 50-item questionnaire took 6.3 (4.9) min to complete, and 142 (89%) patients completed it within 10 min. Most patients completed the questionnaire without any assistance, although some required prompting if items were overlooked. Mean preoperative QoR-40 was 183 (SD 17) and after operation, 167 (23). Changes in perioperative health status and responsiveness are summarized in Table 3.

The correlation between the postoperative QoR-40 and VAS was r=0.68 ($P<0.0005$). This correlation was consistent for patients recovering from day-case surgery (r=0.76), minor surgery (r=0.66) and major surgery (r=0.72) (all $P<0.0005$). The correlation between the postoperative QoR-40 and VAS was significantly stronger than that between the VAS and QoR score (0.68 *vs* 0.62) ($P<0.0005$).

Men had a higher QoR-40 score than women (173 (17) compared with 162 (26)) (P=0.002). There was a significant negative correlation between the QoR-40 and duration of hospital stay (rho=–0.24, $P<0.001$), but not for duration of stay in the recovery room (rho=–0.12, P=0.11). There was a negative correlation between the QoR-40 and time required to complete the questionnaire (rho=–0.22, $P<0.001$).

Test–retest reliability and internal consistency for the QoR-40 were high (r_i=0.92 and α=0.93, respectively) ($P<0.0005$). The split-half coefficient was 0.83 ($P<0.0005$). The test–retest bias was +7.1 and the repeatability coefficient was 24. Test–retest reliability was significantly higher for the QoR-40 than for the QoR score (0.92 *vs* 0.71) ($P<0.0005$).

The median item-to-own dimension coefficients and Cronbach's α for each dimension were: emotional state (r=0.66, α=0.82), patient comfort (r=0.63, α=0.83), psychological support (r=0.67, α=0.80), physical independence (r=0.74, α=0.80) and pain (r=0.63, α=0.77). The inter-dimension correlation matrix is shown in Table 4.

Discussion

We have developed and evaluated a 40-item quality of recovery score (QoR-40) in a diverse group of patients recovering from many types of surgery. The validity, reliability and clinical acceptability of the score was excellent, with most patients able to complete the original 50-item questionnaire in less than 10 min.

Because there is no gold standard measurement of good quality postoperative recovery, we chose to validate the QoR-40 using a variety of end-points. Content validity has been demonstrated previously.[14] The evidence of construct validity was strong, with the QoR-40 being able to discriminate between men and women; it is known that women generally have a worse postoperative recovery.[6–9] We were able to demonstrate a negative association between the QoR score and duration of hospital stay. The finding that inter-dimension correlations were moderate, and that each dimension was better correlated with the overall QoR-40, supports construct validity. Reliability was also confirmed. The reliability coefficients of the QoR-40 exceeded published recommendations (>0.70 to 0.80), [4 17] indicating that the QoR-40 should provide reliable assessment for both group and individual measurements and/or comparisons.

We derived five clinically relevant dimensions which

Table 3 Change in health status of patients interviewed before operation (preoperative baseline) and again on the day after surgery (postoperative) (mean (SD)). The QoR score is a nine-item score[14] and the QoR-40 is a 40-item score consisting of five dimensions. Standardized response mean=mean change in score divided by its SD

Score	Maximum possible score	Preoperative	Postoperative	Mean change (95% CI)	% Change from baseline	Standardized response mean
QoR score	18	16.2 (1.9)	13.8 (3.3)	–2.4 (–3.0 to –1.8)	15%	0.73
QoR-40 dimensions						
Emotional state	45	41.6 (9.0)	38.1 (6.0)	–3.5 (–5.3 to –1.7)	8%	0.35
Physical comfort	60	55.4 (9.9)	48.7 (8.5)	–6.8 (–8.7 to –4.8)	13%	0.80
Psychological support	35	34.8 (6.9)	33.1 (3.2)	–1.7 (–3.0 to –0.4)	5%	0.24
Physical independence	25	23.6 (2.5)	18.8 (5.1)	–4.8 (–5.9 to –3.9)	20%	0.91
Pain	35	31.9 (3.9)	28.8 (5.4)	–3.1 (–4.0 to –2.1)	10%	0.59
Global QoR-40	200	181 (17)	167 (23)	–16 (–20 to –11)	9%	0.65

Table 4 Inter-dimension correlations for the QoR-40 (inter-dimension $\alpha=0.85$). Each dimension consists of a number of items (as indicated in parentheses)

	QoR-40	1	2	3	4	5
QoR-40 dimensions						
(1) Emotional state ($n=9$)	0.90	–				
(2) Physical comfort ($n=12$)	0.90	0.79	–			
(3) Psychological support ($n=7$)	0.69	0.57	0.50	–		
(4) Physical independence ($n=5$)	0.69	0.48	0.49	0.44	–	
(5) Pain ($n=7$)	0.81	0.70	0.64	0.47	0.43	–

encompass most aspects of good quality postoperative recovery. Each was internally consistent and correlated well with the global QoR-40. Our psychometric evaluation would support combining of scores into a single index, as has been suggested previously.[1] Gill and Feinstein also suggested 10 criteria by which to judge health status measurements;[1] the QoR-40 can satisfy all of these. Other methods have been recommended when describing the ability of a health status instrument to detect a clinically important treatment effect or responsiveness.[15–17] We calculated standardized response mean values of 0.24–0.91, suggesting a strong ability to detect a clinically important change in quality of recovery, even for small numbers of patients.[16] As might be expected, pain, physical comfort and physical independence were most affected by surgery and anaesthesia in our study. Although dimension scores for emotional state and psychological support decreased after surgery, they were probably already compromised before operation and should be less prone to change if hospital staff are attentive to patients' psychological wellbeing throughout the perioperative period. It is likely that scores would have been lower if patients were interviewed at an earlier time after surgery (most would have already recovered significantly from anaesthesia), and this would have detected greater changes in health status and responsiveness.

Similarly, some of the validity and reliability coefficients may be underestimates in view of the nature of our study population. Many of our patients were anxious before operation and some were medically unstable or in pain; these circumstances probably did not provide an ideal baseline for comparison: a QoR-40 measurement after complete recovery may have achieved higher estimates. The test–retest coefficient is also likely to be an underestimate in view of the general ongoing improvement in patients' health status after operation. This contention is supported by the positive test–retest bias.

Patient priorities may differ from those of anaesthetists and surgeons. Postoperative recovery has traditionally been evaluated using a variety of physical, psychological and economic end-points.[5–10] In some circumstances it may be more relevant to measure patients' quality of recovery[14] or their satisfaction with care.[5 12 13] Emphasis on quality becomes more important given the safety of modern anaesthesia.[6 7 9] If most patients can be anaesthetized safely and recover after surgery, then efforts should be made to improve their quality of recovery. Any dimension of the QoR-40 can be targeted at specific assessment if a patient rates it as particularly important to them.[1 2] For example, a patient may wish to regain their usual physical functioning as soon as possible or avoid all pain and physical discomfort, whereas another may require specific emotional and psychological support. With increased interest in quality of care, contemporary hospital practice focuses more strongly on individual patient requirements and thus should place greater emphasis on patient-rated assessment of care.

We have reported previously the validity and reliability of a shorter nine-item instrument, the QoR score.[14] Our present study has confirmed its psychometric properties and demonstrated its responsiveness, but we also found that the QoR-40 had superior validity and reliability. The QoR-40, with its relevant dimensions, should also be more discriminative. This suggests that the QoR score should be reserved for conditions where a simple, rapid evaluation is required (such as audit or other quality assurance activity). The QoR-40 provides a more extensive, yet efficient evaluation of a patient's quality of recovery after anaesthesia and surgery. The QoR-40 would be a useful outcome measure in perioperative clinical studies and for assessing the impact of changes in health care delivery on quality of care.

Acknowledgement

Earlier work, which formed the basis of this study, was funded by a research grant from the Australian and New Zealand College of Anaesthetists.

References

1 Gill TM, Feinstein AR. A critical appraisal of the quality of quality-of-life measurements. *JAMA* 1994; **272**: 619–26

2 Guyatt GH, Cook DJ. Health status, quality of life, and the individual. *JAMA* 1994; **272**: 630–1

3 Editorial. Quality of life and clinical trials. *Lancet* 1995; **346**: 1–2

4 McDowell I, Newell C. *Measuring Health: a Guide to Rating Scales and Questionnaires*. New York: Oxford University Press, 1987

5 Lee A, Lum ME. Measuring anaesthetic outcomes. *Anaesth Intensive Care* 1996; **24**: 685–93

6 Cohen MM, Duncan DG, Pope WDB, Wolkenstein C. A survey of 112,000 anaesthetics at one teaching hospital (1975–1983). *Can J Anaesth* 1987; **33**: 22–31

7 Forrest JB, Cahalan MK, Rehder K, *et al*. Multicenter study of general anesthesia. II. Results. *Anesthesiology* 1990; **72**: 262–8

8 Cohen MM, Duncan PG, DeBoer DP, Tweed WA. The postoperative interview: assessing risk factors for nausea and vomiting. *Anesth Analg* 1994; **78**: 7–16

9 Myles PS, Hunt JO, Moloney JT. Postoperative 'minor' complications: comparison between men and women. *Anaesthesia* 1997; **52**: 300–6

10 Warden JC, Borton CL, Horan BF. Mortality associated with anaesthesia in New South Wales, 1984–1990. *Med J Aust* 1994; **161**: 585–93

11 Moerman N, van Dam FSAM, Oosting J. Recollections of general anaesthesia: a survey of anaesthesiological practice. *Acta Anaesthesiol Scand* 1992; **36**: 767–71

12 Tong D, Chung F, Wong D. Predictive factors in global and anesthesia satisfaction in ambulatory surgical patients. *Anesthesiology* 1997; **87**: 856–64

13 Dexter F, Aker J, Wright WA. Development of a measure of patient satisfaction with monitored anesthesia care. *Anesthesiology* 1997; **87**: 865–73

14 Myles PS, Hunt JO, Nightingale CE, *et al*. Development and psychometric testing of a quality of recovery score after general anesthesia and surgery in adults. *Anesth Analg* 1999; **88** 83–90

15 Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987; **40**: 171–8

16 Katz JN, Larson MG, Phillips CB, Fossell AH, Liang MH. Comparative measurement sensitivity of short and longer health status instruments. *Med Care* 1992; **30**: 917–25

17 Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis* 1985; **38**: 27–36

18 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **i**: 307–10