Validation of the Swedish version of Quality of Recovery score - 15: a multicentre, cohort study

S. Lyckner[1], I.-L. Böregård[2], E.-L. Zetterlund[3] and M. S. Chew[3]

[1]Department of Anaesthesia, Mälarsjukhuset, Eskilstuna, Sweden
[2]Department of Anaesthesia, Södertälje Hospital AB, Södertälje, Sweden
[3]Department of Medical and Health Sciences, Department of Anaesthesia and Intensive Care, Linköping University, Linköping, Sweden

Correspondence
S. Lyckner, Department of Anaesthesia, Mälarsjukhuset, Eskilstuna, SE-631 88, Sweden
E-mail: sara.lyckner@dll.se

Conflict of interest
No conflicts of interest are declared.

Funding
This work was financially supported by grants from Sinnescentrum, Östergötland County Council, Sweden and Centre for Clinical Research Sörmland, Uppsala University, Eskilstuna, Sweden

Sara Lyckner and Ida Linnea Böregård contributed equally to this paper.

Submitted 15 January 2018; accepted 17 January 2018; submission 11 October 2017.

Citation
Lyckner S, Böregård I-L, Zetterlund E-L, Chew MS. Validation of the Swedish version of Quality of Recovery score -15: a multicentre, cohort study. Acta Anaesthesiologica Scandinavica 2018

doi: 10.1111/aas.13086

Background: Quality of recovery (QoR) after surgery is often focused on morbidity, mortality and physiological changes, while well-being and emotional state are other important aspects that are often ignored. QoR is poorly investigated in clinical settings and a psychometrically tested questionnaire, QoR-15, has recently been developed. QoR-15 has not been validated for Swedish conditions. The aim of this study was to translate, adapt and validate QoR-15 to Swedish conditions (QoR-15swe).
Methods: A translation and cultural adaption was performed resulting in a Swedish version of the instrument, QoR-15swe. Patients answered the QoR-15swe before surgery, 24 and 48 h after surgery. Feasibility, validity, reliability and responsiveness of the QoR-15swe were evaluated.
Results: The QoR-15swe was feasible in 85.5% of the eligible patients. Construct validity was good, with significant correlations between QoR-15swe score and, ASA-PS class, grade of surgery, length of surgery and time in the post-anaesthesia care unit. The instrument demonstrated good internal consistency with an inter-item Cronbach's α of 0.83–0.87, and inter-dimension Cronbach's α was acceptable 0.71–0.76. Test-retest repeatability was also good with Cronbach's alpha > 0.99 and an interclass correlation coefficient of 0.992 (CI: 0.981–0.997). There were no floor and ceiling effects. Responsiveness assessed by Cliff's effect size was −0.23 indicating a moderate ability to detect change at 24 h postoperatively.
Conclusion: We have translated and culturally adapted the QoR-15 into Swedish. The score demonstrated acceptable validity, reliability and responsiveness. The QoR-15swe is a clinically acceptable and feasible outcome measure after surgery in a Swedish population.

Editorial comment
The authors evaluated a Swedish version of a questionnaire that is used for assessment of quality of recovery after surgery, the QoR-15. They found that it could be administered in 3 min and it had adequate properties, such as the expected correlation with comorbidity, and type as well as duration of surgery.

acta Anaesthesiologica Scandinavica
An international journal of anaesthesiology, intensive care, pain, and critical emergency medicine
ssai

In Sweden, approximately 700,000 surgical procedures are carried out yearly and are perceived as safe treatments.[1] Outcome after surgery is often focused on morbidity, mortality and physiological changes. However, quality of recovery (QoR) after surgery encompasses much more than physiological and physical recovery, and is dependent also on the patients' perspectives on mental well-being including emotional state and perceived psychological support. Allvin et al.[2] identified that the essence of recovery after surgery means to regain mental and physical health. To capture quality of recovery from the patients' perspective, a variety of QoR instruments have been developed.[3–6]

One of them, the 40-item QoR-40, has been the most extensively validated and has demonstrated excellent psychometric properties. It has been translated and validated in different languages[7,8] Although the instrument is well-validated, the feasibility of administering a 40-item questionnaire was questioned by Stark et al.[4]

Quality of recovery 15 (QoR-15) was developed to assess the recovery in a more simplified and user-friendly way without reducing the quality of the instrument.[4] QoR-15 was based on the original instrument and performed well in all dimensions of postoperative recovery. In its original evaluation, QoR-15 was found to be a valid score with good validity, reliability, responsiveness, and feasibility.[4]

QoR-15 is a unidimensional measurement of quality of recovery measuring in five domains: pain, physical well-being, physical independence, psychological support, and emotional state. The instrument is based on 15 questions that results in a numerical score of 0–150 where high level indicates good quality of recovery.[4] It was first translated and validated in Danish,[9] and has recently been translated and validated in Chinese and Portuguese.[10,11] There is no validated translation of the short-form version into Swedish conditions. The more languages and cultural adaptations the instrument is modified to, the greater the possibility of studying and comparing larger populations at international level.[12]

The aim of this study was to translate, adapt and validate QoR-15 to Swedish conditions. We also aimed to report our results according to the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) initiative[13] using the modified criteria by Terwee et al.[14] Our hypothesis was that the Swedish instrument demonstrates comparable validity, reliability, responsiveness and feasibility as the original version in English.[4]

Methods

Translation and cultural adaption

The forward and back translation method was used according to WHO's Process of translation and adaptation of instruments.[12,15] First, the English-language instrument was translated forward to Swedish by two independent bilingual and bicultural translators with Swedish as a mother tongue. One of these translators was experienced with health care terminology. These two translations were compared and merged to create a unified pilot Swedish instrument. The pilot instrument was then blindly back-translated to English by two independent translators. We resolved ambiguities by involving all four translators, and the authors of the present paper. This pre-final version of the Swedish instrument, QoR-15swe [pilot] was then tested in patients and an expert group consisting of nurses with experience in postoperative care.

We a priori defined that if the instrument was rated as 'unclear' by more than 20% of participants then the questions should be reformulated, and the whole test re-evaluated. The minimum inter-rater agreement in both groups was set at 80%. A final Swedish version (QoR-15swe) was determined after fulfilling these criteria.

Inclusion

After approval by the Regional Ethical Review Board, Linköping University, Linköping, Sweden on 26 October 2017 (no. 2016/235-31), psychometric evaluation was conducted in 180 adult (≥ 18 years old) patients undergoing abdominal, breast, orthopaedic, gynaecology and head and neck procedures under general anaesthesia at Linköping University Hospital, Mälarsjukhuset Eskilstuna or Södertälje Hospital in Sweden. To be included in the study

patients had to provide informed consent, understand, speak and read in Swedish, be available for baseline measurements preoperatively and follow-up at 24 and 48 h after surgery either in hospital or via telephone. For logistical and staffing reasons, only patients admitted to PACU between 07:00 and 17:00 hours on weekdays were considered for the study. We attempted to reduce selection bias by allocating 10 random days within this inclusion period for data collection at each hospital.

Data collection

Study nurses at each study site interviewed patients at the bedside during their hospital stay or by telephone if needed. We collected the following data: age, gender, ASA-Physical Status Classification Score (ASA-PS), comorbidities (cardiac ischemia, heart failure, other cardiovascular disease, liver cirrhosis, asthma/COPD, diabetes mellitus, malignancy with metastasis), length of surgery, type of surgery, (elective or acute surgery), grade of surgery according to the Surgery Outcome Risk Tool classification (SORT)[16] and time at post-anaesthesia care unit (PACU). QoR-15swe was measured at three time points – preoperatively, 24 and 48 h after surgery.

Psychometric evaluation

The measurement properties of QoR-15swe was evaluated according to the three domains described by Consensus based Standards for the selection of health Measurement Instruments (COSMIN) initiative.[13] Each domain was rated as positive, negative or indeterminate according to a modified 'Criteria for good measurement properties' described in Prinsen et al.[17] The domains that we evaluated were:

1. Construct validity, assessed as the relationship between QoR-15swe score and age, gender, length of surgery, grade of surgery (minor-intermediate, major, extra-major/complex according to the SORT classification[16]), length of stay in PACU and length of hospital stay. Cross-cultural validity was assessed by comparison with the original version in English.[4]
2. Reliability, assessed as internal consistency measured as inter-item and inter-dimension correlations and Cronbach's alpha. We also assessed test-retest reliability using the inter-class correlation coefficient and Cronbach's alpha. The latter was tested between two repeated measurements in a convenience sample of 24 patients 30–60 min after initial test.
3. Responsiveness assessed as Cliff's effect size as well as using the Kruskal-Wallis test for differences between preoperative and postoperative measurements (24 and 48 h after surgery). We also assessed floor and ceiling effects, defined to be present if greater than 15% of subjects achieve the highest or lowest possible scores. The median difference in QoR-15swe scores was calculated between baseline and 24 h, and between 24 and 48 h post-surgery to assess whether the minimally clinically important difference as defined by Myles et al.[18] was reached.

To evaluate feasibility, we calculated the number of complete responses as a proportion of the total number of included patients (recruitment rate) and average time to successful completion of the questionnaire.

Statistics

Using Souza et al.[12] as guidance, we chose a sample size of 10 patients for each item in the questionnaire (150 patients). We increased the sample size to 180 patients to account for missing data. This sample size is consistent with that in the original study by Stark et al[4] as well as a Danish validation study.[9]

Descriptive statistics were used to characterize the population and reported as mean with standard deviation (SD) or as median with interquartile range (IQR). Associations were measured using the Spearman rank correlation coefficient (ρ). Internal consistency measured using Cronbach's alpha where a score of ≥ 0.9 indicated excellent, 0.8–0.89 good, 0.7–0.79 acceptable and < 0.7 as questionable/poor internal consistency. For differences between groups we used the Kruskal-Wallis test, with a post hoc analysis using the Mann-Whitney test. Responsiveness was measured as Cliff's effect

size due to the presence of categorical variables. Cliff's effect sizes of 0.15, 0.33 and 0.47 correspond to small, medium and large changes in QoR scores (corresponding Cohen's effect sizes of 0.2, 0.5, and 0.8)0.[19] We also assessed responsiveness by calculating the QoR-15swe scores at each time point, and assessing for changes over time using the Friedman's test. A post hoc Wilcoxon test was applied to isolate the groups that were statistically significant from each other.

Only patients completing the questionnaire at all three time points were included in the statistical analysis. A two-sided $P < 0.05$ was considered statistically significant.

The statistical analyses were carried out using IBM SPSS statistics version 22.0.0.0 (SPSS Inc., Chicago, IL, USA).

A copy of the STROBE statement for this study is available in Table S1.

Results

Translation and cultural adaption

After the forward and back translation of the original instrument in English to Swedish the Swedish version of QoR-15 (QoR-15swe[pilot]) was tested. Thirty patients and an expert group of eleven nurses participated. We asked participants if each question and possible responses were understandable and clear. We compared responses within each group, and between groups. An inter-rater agreement $\geq 80\%$ was considered acceptable. A minor re-formulation of language in question seven was made. In the directly translated version of this question we change the asked question 'have you received support from doctors and nurses?' to 'have you received support from hospital staff?' to reflect the team approach to postoperative care that is ubiquitous in Sweden. We also changed the wording of question 11, 'moderate pain' to 'moderately severe pain' which is more consistent with Swedish expression. These changes resulted in a final Swedish language version, the QoR-15swe.

Feasibility

A total of 665 patients were screened for the study between November 2016 and January 2017. Of these, 214 did not meet the inclusion criteria, 253 were not considered for the study since the study nurse was not available on the random days chosen for the study at each site, 18 patients declined participation, leaving 180 eligible patients for the study. Of these, a further 26 were excluded due to incomplete data. The final sample consisted of 154 patients (Fig. 1). Patient characteristics are shown in Table 1. The QoR-15swe was feasible in 154 of 180 (85.5%) eligible patients.

The average time taken to complete the questionnaire in a sample of 24 patients was 3.0 ± 0.7 min.

Validity

Construct validity

There were significant correlations between surgical grade and QoR-15swe score postoperatively ($\rho = -0.395$, $P = < 0.001$ (24 h) and $\rho = -0.306$, $P = < 0.001$ (48 h)). A Kruskal-Wallis test confirmed the relationship between the grade of surgery and QoR-15swe score at 24 h (minor+intermediate 129[112–134] vs. major 114 [102–131] vs. extra-major/complex 104[78–113], $P < 0.001$). This relationship was also seen at

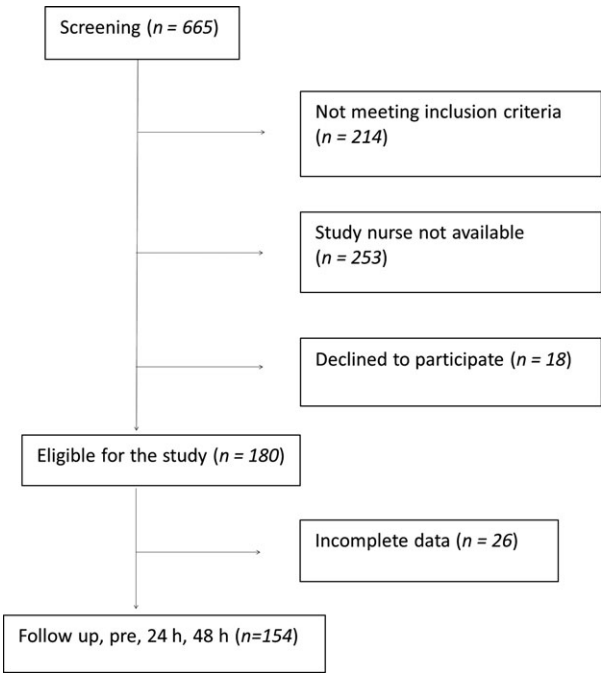

Fig. 1. Study flow chart of the patient recruitment.

Table 1 Patient demographic and surgical characteristics (n = 154).

Age (years)	
Mean ± SD	57 ± 16
Male	50 (33)
Hospital	
Linköping	56 (36)
Mälarsjukhuset	48 (31)
Södertälje	50 (33)
Day surgery	73 (47.4)
Elective procedures	150 (97.4)
Surgical grade	
Minor	6 (3.9)
Intermediate	54 (35)
Major	50 (32.5)
Extra-major/complex	44 (28.6)
ASA-PS class	
I	57 (37)
II	66 (43)
III	31 (20)
IV	0
Presence of at least one comorbidity	57 (37)
Duration of surgery (min)	66 [40–95]
PACU time (min)	158.5 [100–243]
Hospital LoS (days)	1.5 [1–3]

Data are presented as absolute number (%) or median [IQR], unless otherwise stated. LoS, length of stay; PACU, Postoperative Care Unit.

48 h (minor+intermediate 134[122–140] vs. major 128[108–137] vs. complex/extra-major 112 [88–112], $P < 0.0019$). There were no differences due to age or gender. There were significant correlations between ASA-PS class and QoR-15swe scores at all timepoints ($\rho = -0.176$, $P = 0.029$ (pre-score), $\rho = -0.167$, $P = 0.039$ (24 h), $\rho = -0.175$, $P = 0.03$ (48 h). The relationship between ASA-PS class and QoR-15swe was confirmed with a Kruskal–Wallis test ($P = 0.048$). Similar relationships were seen with length of surgery ($\rho = -0.274$, $P < 0.001$ for 24 h, $\rho = -0.294$ $P < 0.001$ for 48 h postoperatively), and length of hospital stay ($\rho = -0.181$, $P = 0.024$ for 24 h, $\rho = -0.399$, $P < 0.001$ for 24 h, $\rho = -0.382$, $P < 0.001$ for 48 h). A significant negative correlation was seen with PACU stay after surgery ($\rho = -0.305$, $P < 0.001$ for 24 h, $\rho = -0.226$, $P < 0.001$ for 48 h).

Cross-cultural validity

The final translated and cultural adapted Swedish version, QoR-15swe, seemed to be an adequate reflection of the original version in English. The translated instrument was perceived as easy to understand and respond to by health care professionals and patients, and adequately reflected the dimensions of postoperative recovery.

Reliability

The inter-item and inter-dimension correlation matrices for QoR-15 swe at 24 and 48 h postoperatively are shown in Tables 2 and 3. Inter-item Cronbach's alpha was 0.86, 0.87, and 0.83 for QoR-15swe measurements at baseline, 24 and 48 h postoperatively. The corresponding values for inter-dimension Cronbach's alpha are shown in Table 4. Test-retest reliability measured using Cronbach's alpha for 23 patients was > 0.99. The interclass correlation coefficient was 0.992 (CI: 0.981–0.997) for the total QoR-15swe score.

Responsiveness

Responsiveness, measured as Cliff's effect size at 24 and 48 h postoperatively is shown in Table 5. Floor and ceiling effects were not seen. There were differences between the total QoR-15swe pre-score and those measured at 24 h 126 [109–139] vs. 114 [100–131], $P < 0.001$, and between 24 and 48 h, 114 [100–131] vs. 128 [108–138], $P < 0.001$. Boxplots (Fig. 2) show the magnitude and direction of change of the QoR-15swe score. The score decreased at 24 h before returning to baseline values at 48 h ($P < 0.001$). The median difference in QoR-15swe scores between baseline and 24 h was 12 and between 24 and 48 h post-surgery the median difference were 14.

The overall quality and properties of QoR-15swe reported according to the COSMIN taxonomy[13] and Prinsen et al.[17] is given in Table S2.

Discussion

For the first time the QoR-15 is now translated and validated into Swedish conditions. Our results show that this translation has preserved the validity of the original English version, in terms of construct, internal consistency, test-retest repeatability. We found that the instrument

Table 2 Inter-item correlation matrices for QoR-15swe at 24 and 48 h postoperatively.

QoR SCORE	24 h Total	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15
Spearman's rho																
24 h Total	–															
Q1	0.49	–														
Q2	0.66	0.30	–													
Q3	0.67	0.19	0.40	–												
Q4	0.69	0.30	0.36	0.58	–											
Q5	0.69	0.42	0.47	0.40	0.44	–										
Q6	0.47	0.31	0.30	0.27	0.33	0.54	–									
Q7	0.28	0.25	0.15	0.14	0.19	0.24	0.22	–								
Q8	0.70	0.26	0.41	0.43	0.40	0.54	0.26	0.07	–							
Q9	0.58	0.43	0.39	0.23	0.29	0.45	0.40	0.34	0.38	–						
Q10	0.68	0.31	0.43	0.41	0.37	0.52	0.31	0.21	0.44	0.57	–					
Q11	0.59	0.24	0.35	0.34	0.37	0.28	0.29	0.11	0.38	0.21	0.29	–				
Q12	0.60	0.35	0.27	0.32	0.48	0.37	0.27	0.15	0.36	0.26	0.29	0.48	–			
Q13	0.57	0.23	0.45	0.39	0.26	0.38	0.22	−0.02	0.40	0.27	0.26	0.30	0.34	–		
Q14	0.48	0.49	0.25	0.24	0.29	0.26	0.25	0.20	0.15	0.50	0.31	0.25	0.29	0.184	–	
Q15	0.49	0.33	0.30	0.31	0.41	0.29	0.31	0.18	0.11	0.32	0.36	0.20	0.28	0.224	0.567	–
QoR SCORE	48 h Total	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15
Spearman's rho																
48 h Total	–															
Q1	0.48	–														
Q2	0.62	0.41	–													
Q3	0.61	0.28	0.44	–												
Q4	0.64	0.37	0.47	0.62	–											
Q5	0.45	0.30	0.24	0.16	0.22	–										
Q6	0.40	0.45	0.35	0.22	0.36	0.38	–									
Q7	0.25	0.13	0.00	0.01	0.01	0.06	0.15	–								
Q8	0.68	0.22	0.38	0.37	0.36	0.35	0.21	−0.14	–							
Q9	0.55	0.56	0.42	0.32	0.44	0.26	0.36	0.12	0.26	–						
Q10	0.70	0.41	0.45	0.44	0.47	0.27	0.21	0.05	0.43	0.51	–					
Q11	0.68	0.29	0.36	0.34	0.34	0.29	0.21	−0.02	0.53	0.22	0.39	–				
Q12	0.63	0.37	0.44	0.27	0.39	0.37	0.32	−0.07	0.49	0.33	0.38	0.60	–			
Q13	0.052	0.08	0.44	0.34	0.27	0.21	0.14	−0.10	0.42	0.13	0.32	0.33	0.39	–		
Q14	0.47	0.43	0.29	0.23	0.31	0.28	0.23	0.07	0.18	0.59	0.45	0.23	0.32	0.13	–	
Q15	0.54	0.44	0.38	0.37	0.40	0.22	0.19	0.11	0.22	0.56	0.47	0.19	0.33	0.16	0.68	–

QoR, quality of recovery; Q, question.

was moderately responsive 24 h postoperatively, but with a return to baseline at 48 h. This is consistent with the findings of Chazapis et al.[19] but indicates a lower degree of responsiveness when compared to the original instrument by Stark et al.[4]

Taken together, these findings speak for the robustness of the instrument which has now been validated in five different languages.

The feasibility was taken into consideration as recommended by Prinsen et al.[17] and in our study the feasibility was generally good for elective surgical patients. For emergency surgery, we believe that the feasibility was poor although it was not explicitly demonstrated, since only four patients participated in this study. We were often not aware of these patients until they were already on their way to the operating theater, and thus no baseline measurements were possible, making them ineligible for the study. This needs to be taken into consideration for future studies, as we believe that this situation is not unique to Sweden.

Construct validity was assessed by comparing QoR-15swe scores of patients having

Table 3 Inter-dimension correlation matrices for QoR-15swe at 24 and 48 h postoperatively (n = 154).

	Pain24	PhC24	PhI24	PsS24	ES24
Spearman's rho					
Pain 24	–				
Physical Comfort24	0.550	–			
Physical Independence24	0.461	0.655	–		
Psychological Support24	0.278	0.284	0.341	–	
Emotional State24	0.392	0.562	0.504	0.329	–
	Pain48	PhC48	PhI48	PsS48	ES48
Spearman's rho					
Pain 48	–				
Physical Comfort48	0.528	–			
Physical Independence48	0.574	0.506	–		
Psychological Support48	0.006	0.061	−0.042	–	
Emotional State48	0.454	0.594	0.473	0.136	–

QoR, quality of recovery; PhC, physical comfort; PhI, physical indepencence; PsS, psychological support; ES, emotional state.

Table 4 Inter-dimension Cronbach's alfa at baseline, 24 and 48 h postoperatively.

Test	Cronbach's alfa
15 items, baseline	0.860
15 items, 24 h	0.869
15 items, 48 h	0.830
5 dimensions, baseline	0.759
5 dimensions, 24 h	0.777
5 dimensions, 48 h	0.707

Table 5 Responsiveness of QoR-15swe at 24 and 48 h postoperatively measured using Cliff's effect size (n = 154).

	Pre-24 h	Pre-48 h	24–48 h
Breathing	0.03	0.16	0.12
Food	−0.21	0.02	0.24
Rest	−0.19	0.16	0.34
Sleep	−0.14	0.18	0.29
Hygiene	−0.27	−0.14	0.14
Communication	−0.13	−0.05	0.08
Support	0.11	−0.06	−0.18
Return to work	−0.48	−0.41	0.09
Feeling in control	−0.09	0.06	0.15
Well-being	−0.17	−0.02	0.16
Moderate pain	−0.27	−0.17	0.13
Severe pain	−0.11	−0.03	0.10
Nausea vomiting	−0.17	−0.02	0.16
Anxiety	0.21	0.34	0.13
Depressed	0.12	0.17	0.05
QoRScore	−0.23	−0.01	0.23

QoR, quality of recovery; Pre, pre-surgery.

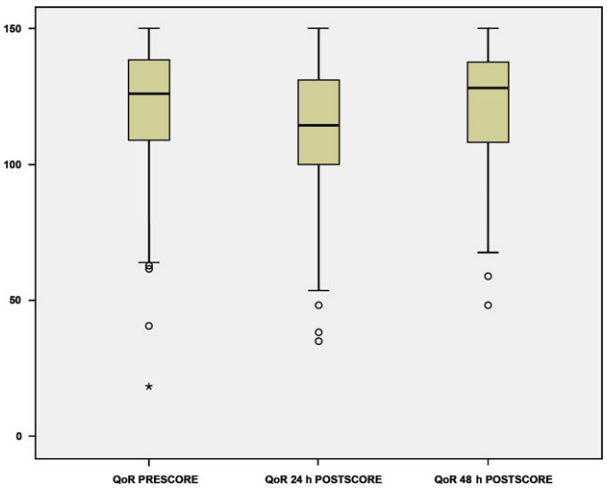

Fig. 2. Boxplots show the magnitude and direction of change of the QoR-15swe score (n = 154). The score decreased at 24 h before returning to baseline values at 48 h ($P < 0.001$). QoR = quality of recovery. Pre-score = scoring before surgery. Post-score 24 h and Post-score 48 h = scoring after surgery at 24 and 48 h, respectively.

minor+intermediate vs. major vs. extra-major/complex surgery where scores for the latter were significantly lower.[20] QoR-15swe scores were significantly correlated with a longer duration of surgery and longer PACU stay, all consistent with previous known factors affecting postoperative recovery.[1,4,19] As for the original instrument and its subsequent validations, we also found a high internal consistency with Cronbach's alpha values ranging from 0.70 to 0.87, indicating good reliability. Interestingly Cronbach's alpha was highest for measurements made 24 h postoperatively and lowest for measurements at 48 h. This suggests that the instrument may be more reliable for measuring recovery at an early stage post-surgery and requires further investigation in different populations (e.g. day surgery vs. emergency surgery vs. major procedures). Our sample size was not

powered to investigate such associations. These results are however slightly discrepant with the findings of Chazapis et.al[19] who found no differences in interdimension correlation at 24 and 48 h. However, even a value of 0.70 may be considered acceptable for scales such as QoR15 that are designed to be used as quick, easy to implement and clinically feasible, and as part of other assessments of perioperative care. We can only speculate as to why our results were slightly different from Chazapis et al.[19] One possible explanation may be that the underlying populations were different from Chazapis et al.[19] where only day surgical patients were enrolled.

Item seven (support from hospital staff) also showed poor inter-item correlations, which is consistent with previous findings.[4,19] We also note that this item became more negative at 48 h postoperatively, which may reflect a perceived lack of support, follow-up or information by patients. Furthermore, there are two possible interpretations of the Swedish translation of item seven: actual support received and perceived support needed by patients. This raises the question of whether the wording of the Swedish version requires revision. However, even the original English item seven is open to interpretation and we note poor inter-item and inter-dimension correlations in previous studies.[4,19]

Test-retest reliability was excellent, in line with previous studies.[4,9]

The responsiveness of QoR-15swe was assessed using Cliff's effect size. We chose this statistic over Cohen's effect size since we believe that the scales used in this questionnaire have ordinal, but not continuous properties. The Cohen's effect size was 1.35 at 24 h postoperatively in the original paper by Stark et al[4] and 0.31 and 0.30 at 24 and 48 h respectively by Chazapis et al.[19] In comparison, our study demonstrated a Cliff's effect size of −0.23 at 24 h and −0.01 at 48 h postoperatively. This translates to only moderate responsiveness at 24 h and poor responsiveness at 48 h which is not consistent with the original paper by Stark et al[4] as well as the Danish translation[9], but in line with the findings of Chazapis.[18,19] One possible explanation could be that recovery occurred much faster in our population, and that QoR-15 scoring should have been made at an earlier stage. This is also reflected by the significant difference between QoR-15swe scores at baseline and 24 h, but not between baseline and 48 h. The median difference in QoR-15swe scores between baseline and 24 h were 12, and between 24 and 48 h post-surgery the median difference were 14, which indicates a clinically meaningful change as defined by Myles et al.[18] The differences between the responsiveness detected in this study compared to Stark[4] and Chazapis[19] may also be related to a difference in study populations where the latter included only day-case surgeries and hence lower severity of surgical trauma and differences in the provision of postoperative care. There was considerable variability between the 15 items, with most responsive items being the ability to return to work or usual home activities (item eight), and the ability to look after personal hygiene (item five). This is also consistent with the results of previous findings.[4,9,19]

In summary, we found the Swedish version of QoR-15 easy to administer with minimal training, with good feasibility in elective surgical patients. Its psychometric characteristics were good, although responsiveness was less than expected compared to the original version of the instrument. As in one previous study,[19] we noted differences in quality of recovery depending on the time of postoperative sampling, and hypothesize that this may be related to different surgical populations, surgical urgency and surgical complexity. A return to baseline values at 48 h postoperatively suggests that multiple testing would be useful to follow the trajectories of recovery in different populations. Two relevant areas of future investigation are the relationships between quality of recovery and postoperative complications as well as health-related quality of life.

Limitations

This study has several important limitations. For logistical and staffing reasons, only patients admitted to PACU between 0700 and 1700 on weekdays were considered for the study. Although we attempted to mitigate this by allocating 10 random days within the inclusion period for data collection at each hospital the

possibility of selection bias remains. Secondly, we found that the vast majority of patients underwent elective procedures. However, this was not due to purposeful exclusion of acute and emergency cases, rather, the latter patients were excluded due to the lack of preoperative QoR measurements since study nurses at the PACU were often not informed of these patients preoperatively. We consider this to be a significant general limitation regarding the feasibility of the QoR-15 instrument.

Although we endeavoured to evaluate QoR-15swe according to the COSMIN taxonomy[13] there were several components of validity that were not included in our study. These were content validity with face validity and criterion validity. Face validity was not assessed because it was already a face-validated instrument in use and we only wanted to translate and test in a different cultural setting. Criterion validity was not assessed because we could not identify a gold standard for comparison. The gold standard would be the QoR-40 which has not been translated to Swedish and is directly related instrument, making comparison inappropriate. Another important limitation is that we only evaluated 24 patients for test-retest reliability, which is arguably a small sample size. However, the small variation in the response suggests that sample was adequate. The retest was conducted after 30–60 min, and although we are aware of the risk of recall bias, this short timeframe was chosen to avoid a change in patients' recovery status.

Conclusion

We have translated and culturally adapted the QoR-15 into Swedish.

The score demonstrated acceptable validity, reliability and responsiveness. The QoR-15swe is a clinically acceptable and feasible outcome measure after surgery in a Swedish population.

Acknowledgements

The authors would like to thank research nurses Lena Sundin and Gunilla Gagnö, Department of Anesthesia and Intensive Care, University Hospital, Linköping, Sweden. We are grateful to Hans-G Eriksson, Centre of Research and Development, Sörmland County Council, Sweden, for statistical advice and calculations.

References

1. Nyckeltal 2016 OpAnIVA SHNAa. Nysamrapport. [www document]. Available at: http://www.nysam.com/report/form (accessed 1 October 2017)
2. Allvin R, Berg K, Idvall E, Nilsson U. Postoperative recovery: a concept analysis. J Adv Nurs 2007; 57: 552–8.
3. Gornall BF, Myles PS, Smith CL, Burke JA, Leslie K, Pereira MJ, Bost JE, Kluivers KB, Nilsson UG, Tanaka Y, Forbes A. Measurement of quality of recovery using the QoR40: a quantitative systematic review. Br J Anaesth 2013; 111: 161–9.
4. Stark PA, Myles PS, Burke JA. Development and psychometric evaluation of a postoperative quality of recovery score: the QoR-15. Anesthesiology 2013; 118: 1332–40.
5. Myles PS, Hunt JO, Nightingale CE, Fletcher H, Beh T, Tanil D, Nagy A, Rubinstein A, Ponsford JL. Development and psychometric testing of a quality of recovery score after general anesthesia and surgery in adults. Anesth Analg 1999; 88: 83–90.
6. Royse CF, Newman S, Chung F, Stygall J, McKay RE, Boldt J, Servin FS, Hurtado I, Hannallah R, Yu B, Wilkinson DJ. Development and feasibility of a scale to assess postoperative recovery: the post-operative quality recovery scale. Anesthesiology 2010; 113: 892–905.
7. Karaman S, Arici S, Dogru S, Karaman T, Tapar H, Kaya Z, Suren M, Gurler Balta M. Validation of the Turkish version of the Quality of Recovery-40 questionnaire. Health Qual Life Outcomes 2014; 12: 8.
8. Tanaka Y, Wakita T, Fukuhara S, Nishiwada M, Inoue S, Kawaguchi M, Furuya H. Validation of the Japanese version of the quality of recovery score QoR-40. J Anesth 2011; 25: 509–15.
9. Kleif J, Edwards HM, Sort R, Vilandt J, Gogenur I. Translation and validation of the Danish version of the postoperative quality of recovery score QoR-15. Acta Anaesthesiol Scand 2015; 59: 912–20.
10. Bu XS, Zhang J, Zuo YX. Validation of the Chinese version of the quality of recovery-15 score and its comparison with the post-operative quality recovery scale. Patient 2016; 9: 251–9.
11. Sa AC, Sousa G, Santos A, Santos C, Abelha FJ. Quality of recovery after anesthesia: validation of the portuguese version of the 'quality of recovery 15' questionnaire. Acta Med Port 2015; 28: 567–74.

12. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. J Eval Clin Pract 2011; 17: 268–74.
13. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010; 63: 737–45.
14. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007; 60: 34–42.
15. Organization WH. Recearch tools: Process of translation and adapation of instruments. [www document]. Available at: http://www.who.int/entity/substance_abuse/research_tools/translation/en/WHO (accessed 1 October 2017).
16. Protopapa KL, Simpson JC, Smith NC, Moonesinghe SR. Development and validation of the Surgical Outcome Risk Tool (SORT). Br J Surg 2014; 101: 1774–83.
17. Prinsen CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, Williamson PR, Terwee CB. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. Trials 2016; 17: 449.
18. Myles PS, Myles DB, Galagher W, Chew C, MacDonald N, Dennis A. Minimal clinically important difference for three quality of recovery scales. Anesthesiology 2016; 125: 39–45.
19. Chazapis M, Walker EM, Rooms MA, Kamming D, Moonesinghe SR. Measuring quality of recovery-15 after day case surgery. Br J Anaesth 2016; 116: 241–8.
20. Vargha A, Delaney HD. A critique and improvement of the CL common language effect size statistics of McGraw and Wong. J Educ Behav Stat 2000; 25: 101–32.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. STROBE Statement—Checklist of items that should be included in reports of *cohort studies.*

Table S2. The overall quality properties of QoR-15swe summarized according to Prinsen et al.[17] and according to the COSMIN taxonomy.[13]