Translation and psychometric evaluation of the Swedish versions of the Nuss Questionnaire modified for Adults and the Single Step Questionnaire

Louise Norlander 1,2*, Jan Karlsson 3, Agneta Anderzén-Carlsson 3, Mårten Vidlund 2, Mats Dreifaldt 2, Jesper Andreasson 4 and Ann-Sofie Sundqvist 2,3

Abstract

Background: Pectus excavatum (PE) is the most common congenital chest wall deformity. Most individuals with PE suffer from psychosocial problems, with low self-esteem and poor body image. Correctional surgery for PE is available, the most widely used is the Nuss procedure. The Nuss procedure has previously been reported to improve self-esteem, body image and health-related quality of life (HRQoL). To assess HRQoL among individuals with PE, the Nuss Questionnaire modified for Adults (NQ-mA) and Single Step Questionnaire (SSQ) has been developed. The aim of the current study was to translate and culturally adapt NQ-mA and SSQ to fit a Swedish context, and to evaluate the psychometric properties, and validate the culturally adapted versions.

Methods: Individuals who had undergone the Nuss procedure for pectus excavatum were invited to participate in a multicentre study with cross-sectional design. HRQoL was assessed by NQ-mA, SSQ and RAND-36. Psychometric properties for NQ-mA and SSQ were evaluated by content validity index and construct validity (exploratory factor analysis, confirmatory factor analysis, and inter-scale correlations). Known-groups validity, as well as floor and ceiling effects, were evaluated. Internal consistency reliability was examined using Cronbach’s alpha.

Results: In total 236 individuals participated in the study. Content validity index for NQ-mA showed satisfactory results, except for two items. Factor analysis for NQ-mA suggested a two-factor model, with exclusion of two items. Subscales correlated adequately with RAND-36’s domains. Known-groups validity for NQ-mA demonstrated high sensitivity for between-group differences. Ceiling effects were present in several items. Cronbach’s alpha for NQ-mA was .89. Confirmatory factor analysis for SSQ resulted in a three-factor model, with exclusion of five items. However, this model was shown to be unstable through further exploratory factor analysis testing, and no further psychometric tests were conducted for SSQ.

Conclusion: The 10-item Swedish version of NQ-mA was shown to be valid for research and clinical assessment of HRQoL in individuals with pectus excavatum. The Swedish version of SSQ requires revision of items before further validation can be performed.

Keywords: Factor analysis, Nuss procedure, Pectus excavatum, Psychometrics, Quality of life
Background
Pectus excavatum (PE) is the most common congenital chest wall deformity, with an incidence of 1 in 400 live births, predominately affecting males [1]. PE is characterized by a depression of the sternum and the adjacent costal cartilages. Physical limitations due to the deformity are primarily cardiopulmonary, such as shortness of breath and exercise incapacity, but other symptoms like chest pain may also occur [2]. However, most affected individuals suffer more from psychosocial distress, with low self-esteem and poor body image [3, 4]. Individuals with PE also report decreased health-related quality of life (HRQoL) compared to individuals without PE [5].

Correctional surgery for PE is available. The most widely used is the minimally invasive technique known as the Nuss procedure. This procedure, performed with thoracoscopic guidance, involves placing one or more convex metal bars beneath the sternum, forcing the depression outward [6]. The bars are removed after 3 years when the chest wall has permanently remoulded [7].

It is frequently assumed that correctional surgery for PE is performed solely for cosmetic reasons. Nevertheless, given that individuals with PE may suffer from psychosocial distress and often experience limitations of social activities, surgery can be life changing [8]. In Sweden, the number of correctional surgeries for PE has increased in recent years. Approximately 72 surgeries were performed annually during 2014–2018 and in 2019, 124 surgeries were performed [9]. It is thus essential to evaluate surgical outcomes adequately, not only from the perspective of physical results but also by the effects on HRQoL [10–13]. Previous international studies confirm that the Nuss procedure improves self-esteem and body image, as well as HRQoL [3, 14, 15]. Instruments that measure disease-specific HRQoL among individuals with PE would therefore be helpful for describing how HRQoL is perceived in a Swedish population.

Several questionnaires have been developed to assess HRQoL in individuals with PE. The Pectus Excavatum Evaluation Questionnaire (PEEQ) was developed by Lawson and colleagues [11] for assessment of the impact of the surgery on the HRQoL of children with PE. PEEQ was further modified by Krasopoulos and colleagues [13] into the Nuss Questionnaire modified for Adults (NQ-mA). Krasopoulos also designed the Single Step Questionnaire (SSQ) [13], in addition to NQ-mA, as a simpler one-step instrument to evaluate satisfaction after surgery. NQ-mA and SSQ are the most frequently used instruments for assessment of HRQoL in adults with PE [5, 10, 12, 13, 15–22].

NQ-mA and SSQ are potentially useful instruments to assess HRQoL in adults who have undergone the Nuss procedure in Sweden. However, none of these versions have previously been validated. In order to ensure that HRQoL-instruments provide valid, accurate and interpretable data, it is important to evaluate their psychometric properties [23]. Therefore, the aim of the current study was to translate and culturally adapt NQ-mA and SSQ to fit a Swedish context and to evaluate their psychometric properties and validate the culturally adapted versions.

Methods
Data collection and sample
Cross-sectional data were collected from the three Swedish cardiothoracic departments where the largest numbers of Nuss procedures are performed (Skåne University Hospital in Lund, Sahlgrenska University Hospital in Gothenburg and Örebro University Hospital). Data were collected during spring 2019, starting in April. All individuals who had undergone the Nuss procedure between January 2000 and April 2019 were invited to participate in the study. Individuals under the age of 15 at the start of the study were excluded. A total of 420 individuals were included, they were asked to fill out NQ-mA, SSQ, RAND-36 and questions about demographic characteristics.

The questionnaires were sent out by mail and could be answered either by returning the paper questionnaires or via an Internet link. Standard format lay-out was used for the paper questionnaires. For the electronic questionnaires, only minor lay-out modifications were made [24]. The data entry of the paper questionnaires was performed manually, and all registered answers were controlled twice. If a respondent had answered in-between two response options, the item was considered missing.

Questionnaires
NQ-mA is a 12-item instrument developed to assess the impact of the Nuss procedure on HRQoL [13]. Each item is scored on a four-point Likert scale (Table 1). NQ-mA consists of two domains measuring aspects of psychosocial functioning (PS; items 1–9) and physical functioning (PH; items 10–12) [11]. All items are also summated to a total score, with higher scores indicating better HRQoL.

SSQ consists of 16 items and was developed for use in addition to NQ-mA at follow-up after surgery. Items 1–7 and 11–15 are scored on a five-point Likert scale, items 8 and 9 are scored on a 10-point visual analogue scale, item 10 is scored on a six-point Likert scale, and item 16 on a three-point scale (Table 2). To calculate the total score, item 8 is subtracted from item 9 and the difference is used as a single score, which is summated with the remaining items [13].
### Table 1 Nuss Questionnaire modified for Adults

| Question stem                                      | Scoring                                                |
|---------------------------------------------------|--------------------------------------------------------|
| 1. Looks in general                               | 4. Very happy 3. Mostly happy 2. Mostly unhappy 1. Very unhappy |
| 2. How chest looks without shirt                  | 4. Very happy 3. Mostly happy 2. Mostly unhappy 1. Very unhappy |
| 3. Spending rest of life as chest looks now       | 4. Very happy 3. Mostly happy 2. Mostly unhappy 1. Very unhappy |
| 4. Other people make fun of you because of your chest looks? | 1. Very often 2. Often 3. Sometimes 4. Never |
| 5. Avoid doing things                              | 1. Very often 2. Often 3. Sometimes 4. Never          |
| 6. Hides chest                                    | 1. Very often 2. Often 3. Sometimes 4. Never          |
| 7. Bothered because of the way chest looks??      | 1. Very often 2. Often 3. Sometimes 4. Never          |
| 8. Feels shy/self-conscious because of chest looks?? | 1. Very often 2. Often 3. Sometimes 4. Never        |
| 9. Feels bad about her/himself                    | 1. Very often 2. Often 3. Sometimes 4. Never          |
| 10. Has trouble exercising                        | 1. Very often 2. Often 3. Sometimes 4. Never          |
| 11. Chest causes shortness of breath              | 1. Very often 2. Often 3. Sometimes 4. Never          |
| 12. Chest is the cause to be tired                | 1. Very often 2. Often 3. Sometimes 4. Never          |

The original version of the Nuss Questionnaire modified for Adults, published by: George Krasopoulos, Michael Dusmet, George Ladas, Peter Goldstraw, Nuss procedure improves the quality of life in young male adults with pectus excavatum deformity, European Journal of Cardio-Thoracic Surgery 2006; 29(1): 1–5, doi:https://doi.org/10.1016/j.ejcts.2005.09.018. Reprinted by permission of Oxford University Press

### Table 2 Single Step Questionnaire

| Question stem                                      | Scoring                                                |
|---------------------------------------------------|--------------------------------------------------------|
| 1. Health in general after the operation          | 5. Much better now 4. Somewhat better 3. About the same |
| 2. Exercise capacity after the operation          | 5. Much better now 4. Somewhat better 3. About the same |
| 3. Extent that chest looks interfere with pre-operative social activity | 5. Extremely 4. Quite a bit 3. Moderately 2. Slightly 1. Not at all |
| 4. Extent that chest looks interfere with post-operative social activity | 5. Not at all 4. Slightly 3. Moderately 2. Quite a bit 1. Extremely |
| 5. Satisfaction with the overall post-operative appearance | 5. Extremely satisfied 4. Very satisfied 3. Satisfied 2. Dissatisfied 1. Very dissatisfied |
| 6. Bothered by the surgical scars                 | 5. Not at all 4. Very slightly 3. Slightly 2. A little bit 1. A lot |
| 7. Impact operation had to social life            | 5. Major improvement 4. Improved 3. No change 2. Worse now 1. A lot worse now |
| 8. Pre-operative self-esteem                      | Score: 1–10                                           |
| 9. Post-operative self-esteem                     | Score: 1–10                                           |
| 10. Pain during hospital stay                      | 5. None 4. Very mild 3. Mild 2. Moderate 1. Severe 0. Very severe |
| 11. Pain interfering with day-to-day activity now (5 months post-operatively) | 5. Not at all 4. Very slightly 3. Slightly 2. A little bit 1. A lot |
| 12. Pain now (5 months post-operative)            | 5. No 4. Occasionally 3. Mild - no pain-killers 2. Mild - pain-killers 1. A lot |
| 13. Conscious about the metallic bar               | 5. Not at all 4. Slightly 3. Moderately 2. Quite a bit 1. Extremely |
| 14. Overall satisfaction with the final result    | 5. Extremely satisfied 4. Very satisfied 3. Satisfied 2. Dissatisfied 1. Very dissatisfied |
| 15. Chest looks different                         | 5. Major improvement 4. Improved 3. No change 2. Worse now 1. A lot worse now |
| 16. Going back, would you have the operation again| 10. Yes 5. Un-sure 0. No                              |

The original version of Single Step Questionnaire, published by: George Krasopoulos, Michael Dusmet, George Ladas, Peter Goldstraw, Nuss procedure improves the quality of life in young male adults with pectus excavatum deformity, European Journal of Cardio-Thoracic Surgery 2006; 29(1): 1–5, doi:https://doi.org/10.1016/j.ejcts.2005.09.018. Reprinted by permission of Oxford University Press
RAND-36 is a generic HRQoL instrument developed to reflect the World Health Organization’s (WHO) definition of health [25]. RAND-36 consists of 36 items grouped into eight domains: physical functioning (PF), role-physical (RP), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role-emotional (RE), and mental health (MH). Domain scores range from 0 to 100, with higher scores indicating better HRQoL.

The participants also filled out a study-specific form eliciting demographic information about sex, age, education, occupation, year of surgery, surgical department, whether bars had been removed, and if so, in what year.

**Translation process**

The translation of NQ-mA and SSQ was conducted in four steps, following modified guidelines from the WHO [26]. The translation process included: (1) a forward translation by a certified translator, and (2) a bilingual expert panel that evaluated the adequacy and vocabulary of the translation. A second expert panel with insight into PE, that is, health professionals and individuals living with the diagnosis, assessed the understanding and relevance of the questions and response options. After the two panels had assessed the translation, the number of response options for item 10 in the SSQ was reduced from six to five to fit with the rest of the questionnaire. The responses for item 14 in SSQ were culturally adapted to fit a Swedish context and translated as ‘very satisfied’, ‘satisfied’, ‘neither satisfied nor dissatisfied’, ‘dissatisfied’, and ‘very dissatisfied’. After this, (3) a back translation to English was carried out by two independent certified translators who were unfamiliar with the original wording of the items. The translation process was iterative, and any uncertainties or problems regarding the items were discussed within the research group of health professionals and researchers with experience of instrument development. The back-translated versions of items and response options were sent to the developer, who approved the translations as well as the change in the response options for SSQ item 10.

The fourth step of the translation process was a pilot test. A systematic sample of 42 individuals (10%) from the total study population were included. Participants were asked to fill out the Swedish versions of NQ-mA and SSQ, as well as an evaluation form with questions regarding their comprehension of the questionnaires and whether they found any of the items inappropriate [27]. A total of 24 participants (57%) responded. It took the participants 5–6 min to fill out each questionnaire. The result of the pilot test showed no frequently occurring difficulties with the questionnaires; thus no further changes were made.

**Usability testing**

The usability of the electronic questionnaires was evaluated. Eight individuals with different technical skills performed a usability test. In addition, they were asked to evaluate whether the software was compatible with different platforms and operating systems. All reported that the software was easy to use and could be run on a computer, tablet, or mobile phone with different operating systems [24].

**Statistical and psychometric methods**

Data were analysed using IBM SPSS Statistics version 25.0 [28] and SAS Software 9.4 [29]. Demographic and clinical characteristics are presented as frequencies and means, along with standard deviations (SD).

**Content validity**

The content validity index (CVI) [30] was calculated to evaluate the adequacy of item content, both for items (I-CVI) and scales (S-CVI) by assessment of relevance. The relevance of items was evaluated by the second expert panel (described above). It consisted of seven people: three of them living with PE, as well as two nurses and two surgeons with experience of individuals undergoing the Nuss procedure. The panel scored relevance on a four-point rating scale with the options: (1) not relevant, (2) somewhat relevant, (3) relevant or (4) very relevant. I-CVI was calculated as the number of experts giving a rating of either 3 or 4 for each item divided by the total number of experts. S-CVI was calculated using the mean of the total I-CVIs for the scale (S-CVI/Ave). An acceptable value for I-CVI is ≥.78, and ≥.90 for S-CVI/Ave.

**Construct validity**

Exploratory factor analysis (EFA) was conducted to assess the underlying factor structure of the instruments. Confirmatory factor analyses (CFA) were used to evaluate the validity of the extracted factor structure found by EFA.

EFA, using principal axis factoring, was conducted to examine the construct validity of the instruments. As part of the EFA, the Kaiser-Meyer-Olkin test was conducted to measure whether data were suited for factor analysis. Bartlett’s test of sphericity was used to assess whether data were multivariate distributed. Squared multiple correlations were used to compute prior communality estimates. Kaiser criterion (eigenvalue >1), interpretability, and the explained variance (≥50%) were used to determine the number of factors to be retained. Items with a minimum loading of .40 [31] were considered to contribute to a given factor. The promax rotation method was used [32].

CFA, using polychoric correlations and robust maximum likelihood estimation [32, 33], was conducted to
test the validity of the factor structure derived by EFA. The following goodness-of-fit indices were used: Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR), Goodness of Fit Index (GFI), Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI) [34].

Item-scale convergent validity was examined by corrected item-total correlations, where \( r \geq .40 \) indicates acceptable correlations [34]. Spearman’s rho was used to assess inter-scale correlations; \( r < .30 \) was considered a weak correlation, \( r = .30–.49 \) as moderate, and \( r \geq .50 \) as strong [35]. The convergent and discriminant validity were also tested at the scale level against the domains in RAND-36. We hypothesized that the NQ-mA PH sub-scale would correlate strongly with the PF and RP domains of RAND-36, and that the PS sub-scale would correlate strongly with VT and SF.

Known-groups validity
NQ-mA’s ability to detect differences between groups was examined with known-groups validity, where specific groups were anticipated to score differently. The subgroups were classified depending on how participants responded to item 14 in SSQ: ‘Overall satisfaction with the final result’ (Table 2). Participants who responded ‘very satisfied’ or ‘satisfied’ (subgroup ‘satisfied’) were anticipated to score significantly higher than participants responding ‘neither satisfied nor dissatisfied’, ‘dissatisfied’, or ‘very dissatisfied’ (subgroup ‘dissatisfied’). Cohen’s \( d \) [35] was calculated to estimate the effect size (ES) of between-group differences. ES was calculated by dividing the mean difference by the pooled standard deviation. ES was considered trivial for \( d < .20 \), small for \( d = .20–.49 \), moderate for \( d = .50–.79 \), and large for \( d \geq .80 \). Between-group differences were also tested using the Mann-Whitney U-test, with a statistical significance level of \( p < .05 \) [34].

Floor and ceiling effects
Floor and ceiling effects were calculated as the proportion of participants scoring at the lowest and highest possible levels. Floor or ceiling effects were considered present if \( \geq 50\% \) of the participants answered the lowest or highest possible score respectively for an item [36] or \( \geq 15\% \) for scales [37].

Reliability
The internal consistency reliability coefficient Cronbach’s alpha (\( \alpha \)) was calculated to determine reliability. An \( \alpha > .70 \) indicates valid reliability at the group level, whereas \( \alpha > .90 \) indicates valid use for individual assessment [34].

Results
A total of 236 individuals responded to the questionnaires, giving a response rate of 56%. Demographics and clinical characteristics are presented in Table 3.

NQ-mA
Content validity
I-CVI for NQ-mA showed variable results: item 3 and 12 had a I-CVI of .57 and .28 respectively, while the remaining items had satisfactory results with I-CVI > .86. The S-CVI/Ave for NQ-mA was .82.

| Table 3 Characteristics and demographics |
|-------------------------|------------------|
| n | 236 |
| Sex, n (%) |
| Male | 194 (82.2) |
| Female | 41 (17.4) |
| Other | 1 (0.4) |
| Age, mean (SD) |
| 25.9 (7.1) |
| Min-max | 15–67 |
| Age at surgery, mean (SD) |
| 19.3 (5.4) |
| Min-max | 10–75 |
| Education, n (%) |
| Compulsory | 50 (21.2) |
| Upper secondary | 116 (49.2) |
| Higher vocational education | 19 (8.1) |
| University | 50 (21.2) |
| Missing | 1 (0.4) |
| Occupation, n (%) |
| Employed | 144 (61.0) |
| Student | 80 (33.9) |
| Job applicant | 8 (3.4) |
| Sick leave | 3 (1.3) |
| Missing | 1 (0.4) |
| Department of surgery, n (%) |
| Gothenburg | 62 (26.3) |
| Lund | 85 (36.0) |
| Örebro | 88 (37.3) |
| Missing | 1 (0.4) |
| Year of surgery, n (%) |
| 2000–2004 | 23 (9.7) |
| 2005–2009 | 40 (16.9) |
| 2010–2014 | 72 (30.5) |
| 2015–2019 | 92 (39.0) |
| Missing | 9 (3.8) |
**Construct validity**

EFA of NQ-mA suggested a two-factor model. The Kaiser-Meyer-Olkin test was .90 and Bartlett’s test of sphericity showed $p < .001$. Items 1–9, which focus on psychosocial functioning, had loadings $\geq .40$ on factor 1, while items 10–12, which measure aspects of physical health had loadings $\geq .40$ on factor 2. CFA was conducted to assess the model fit of the two-factor structure. This 12-item model showed a generally poor fit, indicating that the fit between model and data could be substantially improved (Table 4). Inspection of modification indices showed that the covariance between the residuals for items 2 and 3 caused the largest misfit. Very strong correlation between items 2 and 3 ($r = .84$) was also observed, indicating redundancy. Item 2 was seen as a crucial item covering one of the main difficulties of living with PE, and thus contributing unique information about HRQoL in individuals with PE. Item 3 was therefore removed. Subsequently, a two-factor model with 11 items was tested. The model fit improved significantly, but fit indices indicated that further refinement was needed (Table 4). The covariance between the residuals for items 1 and 2 was the largest cause of misfit, and as item 2 was still considered crucial, item 1 was excluded. The two-factor model for the remaining 10 items showed overall acceptable goodness of fit values (Table 4). A chi-square difference test showed a significant improvement in model fit for the 10-item model. T-values for the factor loadings were significant ($p < .001$) and the standardized factor coefficients ranged between .49 and .92 (Table 5). Since two items had been removed, EFA was performed to test the factor structure of the 10-item, two-factor model. A Kaiser-Meyer-Olkin value of .90 and Bartlett’s test of sphericity of $p < .001$ showed that it met the requirements. Eigenvalues, explained variance, and factor loadings had satisfactory values for the 10-item version (Table 5).

Corrected item-total correlations for the 10-item version resulted in overall acceptable values (Table 5). The psychosocial (PS) and the physical (PH) subscales correlated strongly with the total score of the 10-item NQ-mA (Table 6). All inter-scale correlations between the NQ-mA and RAND-36 scales are presented in Table 6. PH correlated strongly with PF and RP of RAND-36, as hypothesized, as well as with GH. PS had moderate correlations with VT and SF. Moderate correlations were also seen between PS and GH, RE and MH.

**Known-groups validity**

Known-groups analysis showed significant differences between the two subgroups for all scales (Table 7). Cohen’s $d$ showed large ES for PS and the total score for the 10-item NQ-mA, indicating good sensitivity to detect differences between groups. ES for PH was small (ES = .44), demonstrating less sensitivity to between-group differences.

**Floor and ceiling effects**

Floor and ceiling effects are presented in Table 5. Six items (4–6, 8, 10 and 12) had high ceiling effects (> 50%), and also for the PS and PH scales (> 15%).

**Reliability**

Cronbach’s alpha for PS indicated valid use for individual assessment, while PF is valid on the group level (Table 5). Cronbach’s alpha for the total scale was .89, representing nearly acceptable value for individual assessment.

**SSQ**

**Content validity**

CVI for SSQ was satisfactory with I-CVI > .86 for all items and a S-CVI/Ave of .97.

Table 4: Confirmatory factor analysis; Goodness-of-fit indices for Nuss Questionnaire modified for Adults and Single Step Questionnaire

| Fit indices       | NQ-mA, 2 factors | NQ-mA, 2 factors | NQ-mA, 2 factors | SSQ, 1 factor | SSQ, 3 factors | SSQ, 3 factors |
|-------------------|------------------|------------------|------------------|---------------|---------------|---------------|
| Number of items   | 12               | 11               | 10               | 14            | 12            | 10            |
| Items excluded    | –                | 3                | 1, 3             | 13            | 6, 10, 13     | 3, 6, 10, 12, 13 |
| Chi-square/df     | 4.59             | 3.2              | 2.25             | 7.14          | 5.38          | 3.4           |
| RMSEA (90% CI)    | .13 (.11–.15)    | .10 (.08–.12)    | .08 (.05–.10)    | .17 (.15–.18) | .14 (.12–.16) | .10 (.08–.12) |
| SRMR              | .05              | .05              | .04              | .11           | .09           | .06           |
| GFI               | .83              | .90              | .93              | .72           | .84           | .91           |
| CFI               | .90              | .94              | .97              | .66           | .84           | .93           |
| TLI               | .87              | .92              | .96              | .60           | .79           | .91           |

Reference values for acceptable fit: Chi-square/df < 3, RMSEA < .05, SRMR < .08 and GFI, CFI, TLI > .90

CFA = confirmatory factor analysis, df = degrees of freedom, RMSEA = root mean square error of approximation, CI = confidence interval, SRMR = standardized root mean square error, GFI = goodness of fit index, CFI = comparative fit index, TLI = Tucker Lewis index.
Construct validity

As item 8 is subtracted from item 9 to calculate the total score, the difference was used as a single item for the psychometric tests. Item 13 was excluded as 61.4% were missing, since these participants had already removed their bar(s). CFA for SSQ was conducted with a one-factor model of all the remaining 14 items. However, this model generally showed poor fit (Table 4) and EFA was therefore conducted to assess what factor structure was supported by the data.

EFA, performed on the 14-item SSQ, resulted in a four-factor model with acceptable values for the Kaiser-Meyer-Olkin test of .83, and Bartlett’s test of sphericity of $p < .001$. However, item 10 did not load on any factor and was excluded. For the remaining 13 items, EFA suggested a three-factor model. However, item 6 had a weak loading of .31 and so was excluded. EFA of the remaining 12 items again indicated a three-factor model (Table 8). Factor 1 consisted of five items, all focussed on appearance. Factor 2 consisted of four items reflecting aspects of physical health, and factor 3 consisted of three items measuring psychosocial aspects and self-esteem.

CFA of the 12-item three-factor model showed improved goodness-of-fit indices, although the indices were still not acceptable. As modification indices indicated problems with item 3 and 12, these items were excluded, resulting in a 10-item model with fairly acceptable fit. However, further EFA testing of the ten remaining items showed that the three-factor structure was unstable. In addition, the loss of information because of the excluded items was extensive, and it was questionable whether the reduced SSQ was still able to provide reliable information about HRQoL. As a result, no further psychometric testing of SSQ was carried out.

---

### Table 5
Nuss Questionnaire modified for Adults: construct validity, missing, responses, floor/ceiling effects, and reliability

| Factor | Eigenvalue | 1 | 2 | Missing (%) | Floor/Ceiling (%) | Corrected item-total correlation | Cronbach’s $\alpha$ | $\alpha$ if item deleted |
|--------|------------|---|---|------------|-------------------|-------------------------------|---------------------|------------------------|
| PS     | 5.26       | 1.55 | |             | 60.88             | .91                           |                     |                        |
| PH     | 2.14       | 1.09 | |             | 57.48             | .83                           |                     |                        |
| Total  | .93        | .69  | |             | .32               | .63                           |                     | .48                     |

Loadings EFA (CFA)

| PS     | .72 (.77) | .07  | 2.5 | 10.2/24.6 | .71 | .90 |
|--------|          |      |     |            |    |     |
| PH     | .44 (.49) | .11  | 2.1 | 4/85.6    | .47 | .92 |
| Total  | .05       | .76 (.80) | 1.7 | 4.2/55.9 | .58 | .71 |

Items with loadings > .40 were considered to contribute to a factor. $^a$ Number of subjects scoring at the lowest or highest possible scale level. $^b$ $r > .40$ was acceptable. $^c$ $\alpha > .70$ indicates valid use for group assessment, $\alpha > .90$ indicates valid use for individual assessment.

---

### Table 6
Inter-scale correlations between Nuss Questionnaire modified for Adults and RAND-36

| Scale | NQ-mA | RAND-36 |
|-------|-------|---------|
| PS    | .56   | .50b    |
| PH    | .6b   | .50b    |
| Total | .93   | .69     |

Spearman’s rho for correlations; $r < .30$ is considered weak correlation, $r = .30-.49$ moderate, $r \geq .50$ strong. Strong correlations are bolded. Correlations with the same lowercase letter were hypothesised to correlate.

---

### Table 7
Nuss Questionnaire modified for Adults: known-groups analysis

| Scale | Satisfied, $n = 167$ | Dissatisfied, $n = 65$ |
|-------|---------------------|-----------------------|
| PS    | Mean (SD)           | Mean (SD)             |
| PH    | 24.86 (3.28)        | 18.48 (4.86)          |
| Total | 35.02 (4.40)        | 27.21 (5.99)          |

Higher score indicates better health status. $^*$Subgroups classified based on response to item 14, SSQ. $^b$Mann-Whitney U-test for independent group differences. $^c$Significant at level < .05. $^d$ES, Effect Size according to Cohen’s $d$, $d < .20 =$ trivial ES, $d = .20-.49 = $ small, $d = .50-.79 = $ moderate, $d \geq .80 = $ large
The present study aimed to translate and evaluate the psychometric properties of the Swedish versions of NQ-mA and SSQ. As individuals with PE are reportedly suffering from psychosocial and physical limitations due to their chest wall deformity, it is important to adequately measure HRQoL to assess the impact of corrective surgery. The present study found support that a 10-item Swedish version of NQ-mA could be used to assess HRQoL in individuals who have undergone the Nuss procedure. However, the Swedish version of SSQ showed weak construct validity due to the inability to find a stable factor solution.

The results of the psychometric analyses conducted on the Swedish version of NQ-mA showed the 10-item version to be valid for assessment. However, the NQ-mA showed questionable content validity based on CVI results. Lynn [38] suggests removing all items with fair I-CVI or lower ratings in order to achieve content-valid items before distributing the instrument to participants. However, we used a different strategy, and the decision to retain or delete items was instead based on the aggregated results of the psychometric analyses. Thus, items 3 and 12 with fair and low I-CVI, respectively, were retained until psychometric analyses had been evaluated.

NQ-mA item 12 had the highest rate of missing responses, which may imply that participants had difficulty responding to this item. Whether the chest wall issue causes fatigue depends on the extent to which the individual experiences physical limitations due to the deformity. Factor analysis confirmed that item 12 loaded on the PH factor, and as this item was considered to contribute unique information on how PE affects physical HRQoL, it was retained. Another reason for retaining item 12 is that the PH domain consists of only three items, which is the minimum recommended number of items per factor [39].

EFA suggested a two-factor structure of NQ-mA, which is consistent with the results of factor analysis of the Turkish version [40]. CFA could not, however, confirm the model without allowing covariance errors to correlate. In the present study, two items (1 and 3) were excluded, a decision based on both statistical and theoretical grounds [31]. Item 1 correlated strongly with item 2, indicating that these items measure the same aspect of the latent variable and that one item is sufficient. Item 2 was considered important because it measures one of the most frequently reported issues among individuals with PE [14, 41] and thus contributes unique information about HRQoL. Item 1 focuses on overall appearance, not solely regarding the chest wall, which makes the item too general for assessment of disease-specific HRQoL [42]. Item 3 had questionable I-CVI ratings. It may be difficult for participants to respond to a question regarding the future. Furthermore, the correlated error covariances between item 2 and both items 1 and 3 indicate that some other aspect beyond what is measured by the PS domain is influencing the association between these variables [32]. This is an additional reason to exclude items 1 and 3. The suggested 10-item version of NQ-mA is henceforth referred to as NQ-mA-10.

The correlation between the PS and PH subscales of NQ-mA-10 was moderate, suggesting that the scales measure different domains (Table 6). As hypothesized, strong correlations were seen between the PH subscale and RAND-36’s domains PF and RP, providing evidence for convergent validity [34]. PH correlated strongly with the domain GH as well, suggesting that PH measures aspects of general health. Small to moderate associations were seen for PH and the remaining domains, implying discriminant validity. PS had weak correlations with three of the physical health scales (PF, RP and BP), supporting discriminant validity. However, only moderate correlations were seen between PS and the VT and SF domains of RAND-36, which had been hypothesized to correlate strongly. PS also showed moderate associations with RH, RE and MH. This result suggests that the PS scale measures aspects of HRQoL that are not fully consistent with the RAND-36 domains [43]. Thus, moderate correlations between PS and the mental health scales of RAND-36 (VT, SF, RE and MH) indicate that PS measures unique aspects of psychosocial health in individuals with PE that are not captured by a generic HRQoL instrument. Hence, generic and disease-specific instrument should be used in conjunction, as they provide complementary information on HRQoL.

### Table 8: Single Step Questionnaire: exploratory factor analysis

| Factor | 1 | 2 | 3 |
|--------|---|---|---|
| Explained variance, cumulative (%) | 39.07 | 48.87 | 57.5 |
| Loadings | | | |
| Item 5 | .99 | -.09 | -.14 |
| Item 14 | .93 | -.02 | -.07 |
| Item 15 | .82 | -1.16 | -.16 |
| Item 4 | .61 | .15 | -.36 |
| Item 16 | .51 | .12 | .13 |
| Item 12 | -.11 | .81 | -.10 |
| Item 11 | -.02 | .79 | -.22 |
| Item 1 | .09 | .52 | .30 |
| Item 2 | .07 | .51 | .24 |
| Item 3 | -.31 | -.15 | .86 |
| Item 8-9 | .27 | -.07 | .64 |
| Item 7 | .29 | .19 | .50 |

SSQ item 6, 10 and 13 excluded after exploratory factor analysis exploration. Loadings ≥ .40 are bolded and were considered to contribute to a factor.
Ceiling effects were present in several items of NQ-mA-10, and both subscales had ceiling effects well above the acceptable cut-off value. This result can be explained by the cross-sectional design of the study and the fact that HRQoL was measured only postoperatively. In many cases the questionnaire was filled out several years after surgery. The ceiling effects can thus simply demonstrate that participants experience only minor problems because of the positive effects of the surgery. However, a longitudinal study with assessment both before and after surgery is required to accurately evaluate the ceiling effects. Responsiveness can also be evaluated with such a design, which was not possible in the present study.

The Swedish version of SSQ had satisfactory content validity. However, several problems were noted during the factor analysis process, especially concerning item content. Two of the excluded items (3 and 10) ask participants to recall experiences from the past, prior to surgery. Items constructed in this way may be influenced by recall bias as health status memory is often unreliable, especially after long recall periods [34]. In this regard, item 8 is troublesome as well, as it refers to self-esteem prior to surgery. According to Fayers and Machin [34], questions regarding current health status are preferable. Since SSQ is designed to measure HRQoL after surgery, it may be more beneficial to merge item 8 and 9 into one item asking about the difference in self-esteem experienced after surgery.

The wording of SSQ item 12 was considered suboptimal. This item refers to general sensations of pain that are not solely associated with the chest wall. Consequently, the wording is too unspecific to assess diseasespecific HRQoL in individuals with PE. It has not been established whether the same issues regarding item 12 have arisen in the assessment of other SSQ versions. Casamassima and colleagues [12] used a modified version of SSQ to evaluate satisfaction with surgical results after bar removal. Five items (7, 8, 9, 13 and 15) were excluded, but this modified version was never validated. SSQ has been translated into several other languages [10, 17, 18, 20], but without psychometric evaluations. Thus any issues regarding SSQ items in other versions remain unknown.

Despite improved and almost satisfactory goodness-of-fit indices for the 10-item factor model of SSQ, further EFA testing showed that the factor structure was unstable. Additionally, the loss of information due to the excluded items was considered to be excessive. In conclusion, the wording and content of items need extensive revision before further validation can be carried out.

The present study did not assess test-retest reliability, which is a limitation. It is important to evaluate whether SSQ and NQ-mA scores are stable over time and can provide reproducible results. As the original SSQ was validated through test-retest [13], a comparison to assess whether the Swedish version was consistent with the original version would have been possible.

The Swedish version of SSQ could not be validated due to inadequate construct validity. Further reconstruction of items and a new validation study is required before the Swedish SSQ may be available. Additionally, assessment of HRQoL in individuals with PE prior to surgery is needed to further evaluate the Swedish versions of both NQ-mA and SSQ.

Conclusion
The results of the present study show that the Swedish NQ-mA-10 has adequate reliability, sensitivity and validity, indicating that the psychometric properties of this translated and culturally adapted version are sound. The instrument is thus valid for assessment of HRQoL among individuals who have undergone the Nuss procedure for PE in Sweden.
Competing interests
The authors declare that they have no competing interests.

Author details
1Faculty of Medicine and Health, School of Health Sciences, Örebro University, Örebro, Sweden. 2Department of Cardiothoracic and Vascular Surgery, Örebro University Hospital, Örebro, Sweden. 3University Health Care Research Center, Faculty of Medicine and Health, Örebro University, Örebro, Sweden. 4Department of Cardiothoracic Surgery, Skåne University Hospital, Lund University, Lund, Sweden.

Received: 27 October 2020 Accepted: 2 February 2021
Published online: 24 February 2021

References
1. Rea, G., & Sezen, C. B. (2020). Chest wall deformities. In StatPearls. Treasure Island, StatPearls Publishing.
2. Zuidema, W. P., Oosterhuis, J. W. A., Zipper, M. V., van Baren, R., de Lange-De Klerk, E. S. M., van der Heide, S. M., … van Heurn, L. W. E. (2019). Sports activity in adolescents in the Netherlands with a pectus excavatum: the impact of surgery. Journal of Pediatric Surgery, 54, 1671–1674. https://doi.org/10.1016/j.jpedsurg.2018.11.004.
3. Lomholt, J. J., Jacobsen, E. B., Thastrum, M., & Pilegaard, H. (2016). A prospective study on quality of life in youths after pectus excavatum correction. Annals of Cardiothoracic Surgery, 5, 456–465. https://doi.org/10.1016/j.acs.2016.08.002.
4. Kille, S., Müller, A., Steinmann, C., Reingruber, B., Weber, P., & Martin, A. (2012). Self- and social perception of physical appearance in chest wall deformity. Body Image, 9, 246–252. https://doi.org/10.1016/j.bodyim.2012.01.005.
5. Steinmann, C., Kille, S., Mueller, A., Weber, P., Reingruber, B., & Martin, A. (2011). Pectus excavatum and pectus carinatum patients suffer from lower quality of life and impaired body image: A control group comparison of psychological characteristics prior to surgical correction. European Journal of Cardio-Thoracic Surgery, 40, 1138–1145. https://doi.org/10.1016/j.ejcts.2011.02.019.
6. Nuss, D., Kelly Jr., R. E., Croitoru, D. P., & Katz, M. E. (1998). A 10-year review of a minimally invasive technique for the correction of pectus excavatum. Journal of Pediatric Surgery, 33, 545–552. https://doi.org/10.1016/S0022-3468(98)90314-1.
7. Pilegaard, H., & Licht, P. B. (2017). Minimal invasive repair of pectus excavatum and carinatum. Thoracic Surgery Clinics, 27, 123–131. https://doi.org/10.1016/j.tsc.2017.01.005.
8. Fortmann, C., & Petersen, C. (2018). Surgery for deformities of the thoracic wall: No more than strengthening the patient’s self-esteem? European journal of pediatric surgery: Official journal of Austrian Association of Pediatric Surgery [et al] = Zeitschrift fur Kinderchirurgie, 25, 1472–1476. https://doi.org/10.1055/s-0038-1681025.
9. Hennig, M., Kuebler, J. F., Petersen, C., & Metzelder, M. L. (2012). General practitioners’ assessment highlights excellent patient satisfaction following bar removal after Nuss procedure in children and adolescents. European Journal of Pediatric Surgery: Official Journal of Austrian Association of Pediatric Surgery [et al] = Zeitschrift fur Kinderchirurgie, 22, 222–227. https://doi.org/10.1055/s-0033-1308703.
10. Knudsen, M. V., Groesen, K., Pilegaard, H. K., & Laustsen, S. (2015). Surgical correction of pectus carinatum improves perceived body image, mental health and self-esteem. Journal of Pediatric Surgery, 50, 1472–1476. https://doi.org/10.1016/j.jpedsurg.2014.11.048.
11. Makaroww, T. P., Steyn, R. S., & Naidu, B. V. (2011). Prosthesis-free repair of pectus chest deformity. The British Journal of Surgery, 98, 1660–1665. https://doi.org/10.1002/bjs.7596.
12. Metzelder, M. L., Kuebler, J. F., Leonhardt, J., Ure, B. M., & Petersen, C. (2007). Self and parental assessment after minimally invasive repair of pectus excavatum: Lasting satisfaction after bar removal. The Annals of Thoracic Surgery, 83, 1844–1849. https://doi.org/10.1016/j.athoracsur.2006.12.064.
13. Bostanci, K., Ozalper, M. H., Eldem, B., Ozyurtkan, M. O., Isakka, A., Ermerak, N. O., & Yuksel, M. (2013). Quality of life of patients who have undergone the minimally invasive repair of pectus carinatum. European Journal of Cardio-Thoracic Surgery, 43, 122–126. https://doi.org/10.1016/j.ejcts.2014.166.
14. Kuru, P., Dudak, A., Mursaloglu, H., Aktan, H., Oktay, A., & Yuksel, M. (2015). How pulmonary function changes after pectus excavatum correction surgery. Asian Cardiovascular & Thoracic Annals, 23, 945–949. https://doi.org/10.11171/18492315356466.
15. Souza, A. C. Alexandre, N. M. C., & Guiraldello, E. B. (2017). Psychometric properties in instruments evaluation of reliability and validity (Propriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade). Epidemiologia e serviços de saude: revista do Sistema Unico de Saude do Brasil, 26, 649–659. https://doi.org/10.5123/S1679-49742017000300022.
16. Coors, S. J., Gwaltney, C. J., Hays, R. D., Lundy, J. J., Sloan, J. A., Revicki, D. A., … Basch, E. (2009). Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value in Health, 12, 419–429. https://doi.org/10.1111/j.1755-5399.2008.00470.x.
17. Hays, R. D., & Morales, L. S. (2001). The RAND-36 measure of health-related quality of life. Annals of Medicine, 33, 350–357. https://doi.org/10.1080/078537301750002389.
18. WHO (2020) Process of translation and adaptation of instruments: https://www.who.int/substance_abuse/research_tools/translation/en/. Accessed 20 Mar 2020.
19. Steiner, D. L. N. G., & Cairney, J. (2014). Measurement scales: A practical guide to their development and use, (5th ed., ). Oxford: Oxford University Press.
20. Corp, I. (2017). IBM SPSS statistics for windows, (25.0 ed., ). Armonk: IBM Corp.
21. Inc. S1 (2013). SAS, (9a ed., ). Cary: SAS Institute Inc.
22. Politz, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what’s being reported? Critique and recommendations. Research in Nursing & Health, 29, 489–497. https://doi.org/10.1002/nur.20147.
23. Pittuch, K. A. S. (2016). Applied multivariate statistics for the social sciences: Analyses with SAS and IBM’s SPSS, (6th ed., ). New York: Routledge.
24. Brown, T. A. (2006). Confirmatory factor analysis for applied research. New York: Guilford Publications.
25. Holsgado-Tello, F., Moscoso, S., Barbero-Garcia, I., & Vila, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis, with ordinal variables. Quality and Quantity, 44, 153–166. https://doi.org/10.1007/s11135-008-9190-y.
26. Fayers, P. M. (2016). Quality of life: The assessment, analysis, and reporting of patient-reported outcomes, (3rd ed., ). Chichester: Wiley Blackwell.
35. Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155–159. https://doi.org/10.1037/0033-2909.112.1.155.

36. Cappelleri, J. C., Bushmakin, A. G., Gerber, R. A., Leidy, N. K., Sexton, C. C., Lowe, M. R., & Karlson, J. (2009). Psychometric analysis of the three-factor eating questionnaire-R21: Results from a large diverse sample of obese and non-obese participants. International Journal of Obesity, 33, 611–620. https://doi.org/10.1038/ijo.2009.74.

37. Tenwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., … de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology, 60, 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012.

38. Lynn, M. R. (1986). Determination and quantification of content validity. Nursing Research, 35, 382–385.

39. Velicer, W. F., & Fava, J. L. (1998). Affects of variable and subject sampling on factor pattern recovery. Psychological Methods, 3, 231–251. https://doi.org/10.1037/1082-989X.3.2.231.

40.Bahadir, A. T., Kuru, P., Afacan, C., Ermerak, N. O., Bostanci, K., & Yuksel, M. (2015). Validity and reliability of the Turkish version of the nuss questionnaire modified for adults. The Korean Journal of Thoracic and Cardiovascular Surgery, 48, 112–119. https://doi.org/10.5090/kjtcs.2015.4.8.2.112.

41. Ji, Y., Liu, W., Chen, S., Xu, B., Tang, Y., Wang, X., … Cao, L. (2011). Assessment of psychosocial functioning and its risk factors in children with pectus excavatum. Health and Quality of Life Outcomes, 9, 28–28. https://doi.org/10.1186/1477-7525-9-28.

42. Guyatt, G. H., Bombardier, C., & Tugwell, P. X. (1986). Measuring disease-specific quality of life in clinical trials. Canadian Medical Association Journal, 134, 889–895.

43. Lyons-Thomas, J. (2014). Interscale Correlations. In A. C. Michalos (Ed.), Encyclopedia of quality of life and well-being research, (pp. 3352–3353). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_1519.

**Publisher’s Note**
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.