Lack of cross-cultural validity of the Endometriosis Health Profile-30

Nina Julie Verket[1,2], Marit Helen Andersen[3,4], Leiv Sandvik[5], Tom Gunnar Tanbo[1,6], and Erik Qvigstad[1,7]

Abstract

Introduction: The Endometriosis Health Profile-30 is a disease-specific patient-reported outcome measure of health-related quality of life. Cross-cultural validation of the Endometriosis Health Profile-30 has been performed for several translated versions. The aim of this study was to evaluate the measurement properties of a Norwegian version Endometriosis Health Profile-30.

Methods: This study was designed as a cross-sectional anonymous postal questionnaire study. A total of 157 women with endometriosis were included during a period from 2012 to 2013. Women aged 18–45 years were recruited from the Norwegian Endometriosis Association. Principal components analysis with varimax rotation was used to assess construct validity. Short Form-36 was used to determine convergent validity. Cronbach's alpha was used to measure internal consistency. Intraclass correlation coefficients and paired t-tests were used to evaluate test–retest reliability. Floor and ceiling effects were estimated.

Results: Factor analysis resulted in a three and five-factor model for the core and modular questionnaire, respectively. Factor analysis could not support construct validity of the scales self-image and treatment. The Norwegian version Endometriosis Health Profile-30 demonstrated acceptable internal consistency and test–retest reliability, except for the scale relationship with children. Floor effects were observed for the scales self-image (20.1%), work life (33.9%), relationship with children (34.2%), and medical profession (20.5%).

Conclusion: The construct self-image does not seem to be measured appropriately by the Norwegian version Endometriosis Health Profile-30, suggesting a lack of cross-cultural validity of the Endometriosis Health Profile-30. With multinational studies increasing, adequate translation, cross-cultural adaptation, and cross-cultural validation of instruments are essential to ensure equivalence in languages and cultures other than the original.

Introduction

Chronic diseases such as endometriosis can affect health-related quality of life (HRQoL).[1] HRQoL is a multidimensional concept that refers to the patient's general perception of the effect of her disease and treatment on physical, psychological, and social aspects of daily life.[2–4] HRQoL is commonly assessed as a patient-reported outcome, that is, a clinical outcome reported directly by the patient.[3,5] A patient-reported outcome measure (PROM) of HRQoL can be generic, applicable to patients with a variety of conditions, or disease-specific.[6] Disease-specific instruments may detect change in important aspects of certain conditions not accessible by generic

[1]Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[2]Research Centre for Obstetrics and Gynecology, Oslo University Hospital, Oslo, Norway
[3]Institute of Health and Society, University of Oslo, Oslo, Norway
[4]Department of Transplantation Medicine, Oslo University Hospital, Oslo, Norway
[5]Oslo Center for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway
[6]Department of Reproductive Medicine, Oslo University Hospital, Oslo, Norway
[7]Department of Gynecology, Oslo University Hospital, Oslo, Norway

Corresponding author:
Nina Julie Verket, Research Centre for Obstetrics and Gynecology, Oslo University Hospital Ullevål, Postboks 4956 Nydalen, 0424 Oslo, Norway.
Email: ninaverket@gmail.com

instruments.[7] The Endometriosis Health Profile-30 (EHP-30) is a disease-specific PROM of HRQoL consisting of a core and modular questionnaire.[8,9] The original English version was developed in the United Kingdom and first presented in 2001.[8] The items, or questions, were generated from in-depth interviews of 25 patients with endometriosis visiting a gynecology clinic at a large tertiary referral hospital in Oxford.[8]

The EHP-30 is available in many languages. Evaluation of measurement properties, that is, reliability, validity, and responsiveness, has been performed for several of these, however primarily for the core questionnaire.[10–15] With multinational and multicultural studies increasing, adequate translation, cross-cultural adaptation, and cross-cultural validation are essential to ensure equivalence of a PROM in languages and cultures other than the original.[16] The Consensus-based Standards for the selection of health Measurement INstruments (COSMIN) group has developed user-friendly and easily applicable checklists to evaluate the methodological quality of primary studies on measurement properties.[17] According to these checklists, few, if any, of the EHP-30 validation studies have included adequate sample sizes for test–retest reliability analysis.[18] Test–retest reliability is an important aspect of reliability, ensuring that changes detected by an instrument are not random.[3] However, analysis depends on patients being in stable condition. Although endometriosis is sometimes characterized by disease fluctuation, it is also thought to be stable for longer periods of time. Fewer may be in stable condition among patients attending secondary and tertiary referral centers compared with members of patient registries and patient associations.

The aim of this study was to evaluate the measurement properties of the Norwegian version EHP-30 (NO-EHP-30) and thereby its suitability for future use in endometriosis research in Norway or as part of multinational studies.

Methods

Participants, study design, and data collection

Women with endometriosis were recruited from the Norwegian Endometriosis Association. Inclusion criteria were 18–45 years of age and surgically confirmed diagnosis. Cross-sectional data collection was performed from 2012 to 2013. A set of two anonymous postal questionnaires was sent to potential participants. Each questionnaire included questions on background information, NO-EHP-30, and Short form-36 version 2 (SF-36v2).[19] Participants were asked to fill in the second questionnaire 1 month after completing the first questionnaire, for test–retest reliability analysis. A period of 1 month between the test and retest was chosen to minimize memory effects. A period of 1 month was also thought to increase the chances of the respondents being in the same phase of their menstrual cycle, which in turn may be relevant regarding endometriosis complaints and reporting of HRQoL.

Background information

Background information included age, height, and weight. Diagnostic delay was recorded as year receiving diagnosis minus year the participant started having symptoms. Furthermore, a multiple choice question on organs/anatomic locations affected by endometriosis and two open questions inviting free description of previous and present treatment were included. Finally, the participants were asked whether they had experienced dysmenorrhea, pelvic pain, dysuria, and/or dyschezia during the 4 weeks prior to answering the questionnaire.

EHP-30

The responses are based on patient experiences during the 4 weeks prior to answering the questionnaire. The core questionnaire is composed of 30 items grouped into five scales: pain (11 items), control & powerlessness (6 items), emotional well-being (6 items), social support (4 items), and self-image (3 items). The modular questionnaire is composed of 23 items grouped into 6 scales: work life (5 items), relationship with children (2 items), sexual intercourse (5 items), medical profession (4 items), treatment (3 items), and infertility (4 items). The modular questionnaire is characterized by the possibility of responding only to scales which the patient deems relevant to her. All scales can achieve a minimum score of 0, indicating low disability, and a maximum score of 100, indicating high disability. All items of a scale must be answered to be able to calculate a scale score. The only exception is the scale sexual intercourse, where each item may be relevant independently of the other items of the same scale. Thus, the scale score for the scale sexual intercourse is calculated by omitting items which are not relevant.

Translation and cultural adaptation of the Norwegian version EHP-30

The Norwegian language has two distinct written varieties, "bokmål" and "nynorsk."[20] "Bokmål" is the most commonly used variety. The EHP-30 was therefore translated to "bokmål." The translation and cultural adaptation of the NO-EHP-30 was conducted by Oxford outcomes according to recommended guidelines,[21] (Supplementary material 1).

SF-36v2

The Short form-36 is a generic PROM of HRQoL composed of 36 items, one item assessing health change and 35 items assessing eight health concepts representing eight

scales: physical functioning (10 items), role limitations due to physical problems (4 items), bodily pain (2 items), general health perceptions (5 items), vitality (4 items), social functioning (2 items), role limitations due to emotional problems (3 items), and mental health (5 items).[19,22] All scales can achieve a minimum score of 0, indicating worst possible health, and a maximum score of 100, indicating best possible health. QualityMetric Health Outcomes™ Scoring Software 4.5 from OptumInsight Life Sciences, Inc, was used to score SF-36v2.

Sample size calculation

Correlation coefficients play a central role in this study. We used Fisher's z transformation to estimate 95% confidence interval for a correlation coefficient r.[23] The confidence interval for a correlation coefficient r is widest when $r=0.50$. We consider it sufficient with a precision of ±0.10, that is, when the length of the confidence interval for r is at most 0.20.[10] For a correlation coefficient of 0.50 with a sample of 150 patients, this confidence interval will be 0.40–0.60. We therefore decided to include at least 150 women with endometriosis in our study.

Psychometric evaluation and statistical analysis

Construct validity, reliability, and interpretability of the NO-EHP-30 were assessed. We used the taxonomy, terminology, and definitions of measurement properties suggested by the COSMIN study.[24] Hypotheses-testing was specified as assessment of convergent validity where it could be misinterpreted as hypotheses-testing associated with factor analysis. Reliability was specified as test–retest reliability where it was thought to increase clarity. All analyses were performed with IBM SPSS Statistics, version 22.

Construct validity

Structural validity. Exploratory factor analysis was used to assess structural validity.[25] Principal components analysis with varimax rotation was used to identify the different potential components with eigenvalues greater than 1.[26] Items with factor loadings ≥0.40 in a factor were included in the factor.

Hypotheses-testing. SF-36v2 was used for hypotheses-testing to assess convergent validity.[17,27,28] We hypothesized the strongest correlations between EHP-30 pain and SF-36v2 bodily pain, and EHP-30 emotional well-being and SF-36v2 mental health. We further expected a strong correlation between EHP-30 social support and SF-36v2 social functioning, and EHP-30 work life and SF-36v2 role-physical. After obtaining the results of the factor analyses, we hypothesized a strong correlation between

EHP-30 control & powerlessness and SF-36v2 bodily pain, and EHP-30 relationship with children and SF-36v2 role-physical. Associations between scales of the EHP-30 and the SF-36v2 were calculated by Spearman's rho correlation coefficient. There are no widely accepted criteria for defining a strong versus moderate versus weak correlation.[29] Values 0.20–0.39 were considered to indicate weak correlations, values 0.40–0.59 moderate, values 0.60–0.79 strong, and values 0.80–1.00 very strong correlations.

Reliability

Internal consistency. Cronbach's alpha and corrected item-total correlations were used to measure internal consistency. Cronbach's alpha above 0.70 were considered to indicate acceptable internal consistency reliability for group comparisons, and values above 0.90 for individual comparisons.[28] Item-total correlations were corrected for overlap by omitting the item from the parent scale total. Item-total correlations above 0.40 were considered to indicate acceptable internal consistency.[30]

Test–retest reliability. Intraclass correlation coefficients for agreement and paired t-tests were used to measure test–retest reliability. Intraclass correlation coefficients above 0.70 were considered to indicate acceptable reliability for group comparisons, and values above 0.90 for individual measurements over time.[28,31] Significant differences in mean scores ($p<0.05$) were considered to indicate poor reliability. No significant differences in mean scores were considered to indicate acceptable reliability.

Interpretability

Data completeness, mean scores and standard deviations, floor and ceiling effects, and skewness of score distribution were used to describe the distribution of item responses.[17] Floor or ceiling effects were considered present if more than 15% of respondents scored the minimum value of 0 or the maximum value of 100, respectively.[31]

Ethical approval

This study was approved by the Regional Committee for Medical and Health Research Ethics, division south-eastern Norway (trial registration number: 2011/2213/Regional Committee for Medical and Health Research Ethics, division south-eastern Norway B).

Results

Initially, 150 sets of questionnaires were sent to a random sample of members of the Norwegian Endometriosis Association. Of these, 60 questionnaires were successfully completed and returned. Based on this preliminary

Table 1. Basic characteristics of the participants (n = 157).

Age (years), mean ± 1 SD		35.2 ± 6.5
BMI (kg/m²), mean ± 1 SD		24.8 ± 5.2
Diagnostic delay (years), mean ± 1 SD		8.1 ± 6.5
Diagnosis confirmed by surgery (%)		100
	n	%
Organ affected (n = 148)		
Only peritoneum	10	6.8
Ovaries	98	66.2
Bladder	36	24.3
Vagina	28	18.9
Bowels	54	36.5
Previous treatment (n = 146)		
Analgesic	17	11.6
Hormonal	85	58.2
Surgical	122	83.6
Present treatment (n = 138)		
No treatment	45	32.6
Receiving treatment	93	67.4
Analgesic	28	30.1
Hormonal	73	78.5
Awaiting surgery	4	2.9
Pain experienced past 4 weeks		
Dysmenorrhea (n = 135)	97	71.9
Pelvic pain (n = 152)	129	84.9
Dysuria (n = 154)	52	33.8
Dyschezia (n = 155)	83	53.5

BMI: body mass index; SD: standard deviation.

response rate, an additional 225 sets of questionnaires were sent to a second random sample of members of the Norwegian Endometriosis Association not contacted in the first round. In total, 162 of 375 questionnaires were successfully completed and returned. Five of these were from women with endometriosis who reported that their diagnosis had not been confirmed surgically. These were excluded. Among the 157 included respondents, 94 completed and returned a second questionnaire at a later date. Of these, 10 reported change in treatment or starting new treatment since completing the first questionnaire. Excluding these, test–retest reliability of the NO-EHP-30 could be assessed in 84 of the respondents. The median number of days between answering the first and second questionnaire was 34 (range 7–168). Of the 84 respondents, 61 reported either having menstruation when answering both questionnaires or not having menstruation when answering both questionnaires. Of the 84 respondents, 15 reported having menstruation when answering one questionnaire, and not having menstruation when answering the other. The characteristics of the participants are presented in Table 1.

Construct validity

Structural validity. Factor analysis of the 30 items of the core questionnaire suggested three factors, explaining 70.2% of the total variance. The three-factor model resulted in 20 items loading on the hypothesized scales and 10 items loading on alternative scales (Table 2). Factor analysis of the 23 items of the modular questionnaire suggested five factors, explaining 100% of the total variance. The five-factor model resulted in 15 items loading on the hypothesized scales and 8 items loading on alternative scales (Table 3).

Hypotheses-testing. Correlations between scales of the EHP-30 and the SF-36v2 ranged from −0.63 to −0.81 (Table 4). The correlations are negative because the EHP-30 and the SF-36v2 are scored in opposite directions. All hypotheses were confirmed.

Reliability

Internal consistency. Cronbach's alpha ranged from 0.87 to 0.96 for the original scales of the core questionnaire and from 0.78 to 0.94 for the original scales of the modular questionnaire (supplementary material 2). The corrected item-total correlation coefficients ranged from 0.45 (item 23) to 0.91 for the original scales of the core questionnaire and from 0.55 to 0.89 for the original scales of the modular questionnaire.

Test–retest reliability. Intraclass correlation coefficient for test–retest agreement ranged from 0.80 to 0.85 for the scales of the core questionnaire, and from 0.67 to 0.91 for the scales of the modular questionnaire (Table 5). The mean scale scores did not differ significantly between the first and second measurements. Test–retest reliability analysis including only the 61 respondents reporting either having or not having menstruation when answering both questionnaires, did not alter the general findings (data not shown).

Interpretability

The results are presented in Table 6. Data completeness of at least 97.5% was achieved for all EHP-30 scales. The proportion of participants to whom each scale of the modular questionnaire was relevant, varied from 39.4% (the scale infertility) to 87.2% (the scale sexual intercourse). Floor effect was only found for the scale self-image (20.1%) in the core questionnaire, and for the scales work life (33.9%), relationship with children (34.2%), and medical profession (20.5%) in the modular questionnaire. No ceiling effects were observed. Skewness was low for all the scales.

Discussion

Factor analysis suggested a three-factor model for the EHP-30 core questionnaire, in contrast to the original five-factor model. Items of the scales pain and control & powerlessness loaded on the same factor. A similar finding was

Table 2. Factor analysis of the 30 items of the EHP-30 core questionnaire suggesting a three-factor model.

Items of the EHP-30 core questionnaire		Factor 1	Factor 2	Factor 3
1	Been unable to go to social events because of the pain?	0.84		
2	Been unable to do jobs around the home because of the pain?	0.81		
3	Found it difficult to stand because of the pain?	0.77		
4	Found it difficult to sit because of the pain?	0.79		
5	Found it difficult to walk because of the pain?	0.79		
6	Found it difficult to exercise or do the leisure activities you would like to do because of the pain?	0.78		
7	Lost your appetite and/or been unable to eat because of the pain?	0.71		
8	Been unable to sleep properly because of the pain?	0.72		
9	Had to go to bed/lie down because of the pain?	0.80		
10	Been unable to do the things you want to do because of the pain?	0.82		
11	Felt unable to cope with the pain?	0.75		
12	Generally felt unwell?	0.65		
13	Felt frustrated because your symptoms are not getting better?	0.61	0.44	
14	Felt frustrated because you are not able to control your symptoms?	0.60	0.46	
15	Felt unable to forget your symptoms?	0.41	0.57	
16	Felt as though your symptoms are ruling your life?	0.57	0.63	
17	Felt your symptoms are taking away your life?	0.59	0.63	
18	Felt depressed?		0.41	0.70
19	Felt weepy/tearful?		0.40	0.67
20	Felt miserable?	0.61		0.51
21	Had mood swings?			0.77
22	Felt bad tempered or short tempered?			0.80
23	Felt violent or aggressive?			0.65
24	Felt unable to tell people how you feel?		0.68	
25	Felt others do not understand what you are going through?		0.81	
26	Felt as though others think you are moaning?	0.41	0.44	0.40
27	Felt alone?		0.73	
28	Felt frustrated as you cannot always wear the clothes you would choose?		0.62	
29	Felt your appearance has been affected?		0.62	
30	Lacked confidence?		0.48	0.66

EHP-30: Endometriosis Health Profile-30.
Principal components analysis with varimax rotation. Only factor loadings ⩾0.40 are shown.
In the original EHP-30, items 1–11 belong to the scale "pain," items 12–17 to the scale "control & powerlessness," items 18–23 to the scale "emotional well-being," items 24–27 to the scale "social support," and items 28–30 to the scale "self-image."

demonstrated in the original, Portuguese, and French version EHP-30.[9,14,15] As argued by the developers, it is likely that pain has considerable impact on sense of control and powerlessness. In this study, assessment of convergent validity demonstrated strong correlations between each of the EHP-30 scales pain and control & powerlessness and the SF-36v2 scale bodily pain, supporting this interpretation. Strong correlations were also demonstrated between the EHP-30 scales emotional well-being and social support and the corresponding SF-36v2 scales mental health and social functioning. Thus, the findings in this study support construct validity of four of five scales (pain, control & powerlessness, emotional well-being, and social support) of the core questionnaire.

The fifth scale of the core questionnaire, self-image, consists of three items. The first two items concern the effect of endometriosis on choice of clothing and appearance, and the last item concerns the effect of endometriosis on self-confidence. In factor analysis, the first two items loaded on the scale social support, and the last item loaded on the scale emotional well-being. Thus, the construct self-image does not seem to be measured appropriately by the NO-EHP-30. The lack of association between appearance and self-confidence is likely not exclusive to the Norwegian culture. Subtle differences in exploratory factor analysis technique, that is, performed with or without predefinition of five factors for the core questionnaire, may have masked a similar finding in other translated versions.[14,25]

Factor analysis suggested a five-factor model for the EHP-30 modular questionnaire, in contrast to the original six-factor model. Factor analysis of the modular questionnaire has been performed for the original and French version EHP-30.[9,15] In this study, items of the scales work life and relationship with children loaded on the same factor. A similar finding was demonstrated in the original version, but not in the French version.[9,15] These discrepancies may

Table 3. Factor analysis of the 23 items of the EHP-30 modular questionnaire suggesting a five-factor model.

Items of the EHP-30 modular questionnaire		Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
A1	Had to take time off work because of the pain?	0.97				
A2	Been unable to carry out duties at work because of the pain?	0.97				
A3	Felt embarrassed about symptoms at work?	0.51			0.73	
A4	Felt guilty about taking time off work?	0.97				
A5	Felt worried about not being able to do your job?	0.95				
B1	Found it difficult to look after your child/children?	0.97				
B2	Been unable to play with your child/children?	0.97				
C1	Experienced pain during or after intercourse?			0.91		
C2	Felt worried about having intercourse because of the pain?		0.63	0.72		
C3	Avoided intercourse because of the pain?	0.52	0.71			
C4	Felt guilty about not wanting to have intercourse?			0.88		
C5	Felt frustrated because you cannot enjoy intercourse?			0.92		
D1	Felt the doctor(s) you have seen is (are) not doing anything for you?		0.95			
D2	Felt the doctor(s) think it is all in your mind?		0.95			
D3	Felt frustrated at the doctor(s) lack of knowledge about endometriosis?		0.94			
D4	Felt like you are wasting the doctor(s) time?		0.95			
E1	Felt frustrated because treatment is not working?	0.77			0.42	
E2	Found it difficult coping with the side effects of treatment?		−0.90			0.41
E3	Felt annoyed at the amount of treatment you have had to have?		−0.60		0.74	
F1	Felt worried about the possibility of not having children/more children?					0.90
F2	Felt inadequate because you may not/have not been able to have children/more children?	0.56				0.78
F3	Felt depressed at the possibility of not having children/more children?	0.51			0.59	0.46
F4	Felt that the possibility of not conceiving/not being able to conceive has put a strain upon your personal relationship?	0.52			0.77	

EHP-30: Endometriosis Health Profile-30.
Principal components analysis with varimax rotation. Only factor loadings ≥0.40 are shown.
In the original EHP-30, items A1-5 belong to the scale "work life," items B1-2 to the scale "relationship with children," items C1-5 to the scale "sexual intercourse," items D1-4 to the scale "medical profession," items E1-3 to the scale "treatment," and items F1-4 to the scale "infertility."

be due to difference in daily patterns of work life and child care in these three countries. In this study, factor analysis could not support construct validity of the scale treatment. The three items of the scale treatment loaded on three separate factors. A tendency of the first item of the scale treatment to load on a different factor than the two latter items has been demonstrated by factor analysis with larger samples in both the original and French version EHP-30.[9,15]

The NO-EHP-30 demonstrated acceptable test–retest reliability except for the scale relationship with children of the modular questionnaire, which demonstrated an intraclass correlation coefficient of 0.67. Although the time interval between answering the first and second questionnaire likely was long enough to minimize memory effects, it may have allowed changes in the status of the subject.[32] Exclusion of questionnaires from respondents reporting change in treatment or starting new treatment between assessments, probably reduced this effect. Phase of menstruation did not seem to affect the outcome. The scale relationship with children consists of two items. The second item concerns the ability to play with child/children and implies children of younger age. In the case of children of younger age, the score of this scale may depend not only on the health status of the respondent but also on the health status of the child/children. Thus, this particular scale may be less reliable.

This study is the first to evaluate both test–retest reliability and validity of the core questionnaire of the EHP-30 including adequate sample sizes.[18,33] Regarding the modular questionnaire, the varying relevance of scales to participants has likely rendered some sample sizes inadequate. To ensure adequate sample size for the least relevant modular questionnaire scale, the general sample size should have been three times larger. On the other hand, these variations in relevance of the scales of the modular questionnaire, would limit the use of the modular questionnaire in most research settings. Another weakness of this study is the lack of representativeness of

Table 4. Convergent validity. Correlations between some EHP-30 scales and relevant SF-36v2 scales.

EHP-30 scale	SF-36v2 scale	Spearman's rho	Two-tailed test p-value
Core questionnaire			
Pain	Bodily pain	−0.81	<0.001
Control & powerlessness	Bodily pain	−0.73	<0.001
Emotional well-being	Mental health	−0.74	<0.001
Social support	Social functioning	−0.63	<0.001
Modular questionnaire			
Work life	Role-physical	−0.68	<0.001
Relationship with children	Role-physical	−0.75	<0.001

EHP-30: Endometriosis Health Profile-30; SF-36v2: Short Form-36 version 2.

Table 5. Test–retest reliability and intraclass correlation coefficients (ICC) with 95% confidence intervals (CI) for test–retest agreement. Comparison of mean scale scores at time 1 and time 2 with p-values.

EHP-30 scales	n	ICC[a]	95% CI	Mean ± SD Time 1	Mean ± SD Time 2	p-value[b]
Core questionnaire						
Pain	79	0.80	0.71, 0.87	35.5 ± 24.0	35.6 ± 23.2	0.93
Control & powerlessness	80	0.80	0.70, 0.87	49.8 ± 27.2	47.7 ± 27.4	0.27
Emotional well-being	78	0.84	0.75, 0.89	39.0 ± 21.0	39.2 ± 21.5	0.94
Social support	81	0.85	0.78, 0.90	42.3 ± 26.1	43.2 ± 26.6	0.56
Self-image	81	0.80	0.70, 0.86	39.2 ± 29.1	39.7 ± 28.8	0.80
Modular questionnaire						
Work life	63	0.86	0.77, 0.91	26.3 ± 27.4	27.7 ± 26.5	0.46
Relationship with children	42	0.67	0.47, 0.81	28.9 ± 23.2	31.3 ± 24.7	0.43
Sexual intercourse	65	0.91	0.86, 0.95	47.5 ± 30.5	46.8 ± 31.9	0.65
Medical profession	35	0.75	0.56, 0.86	40.4 ± 29.7	37.0 ± 27.9	0.33
Treatment	37	0.71	0.50, 0.84	45.9 ± 27.8	46.8 ± 24.5	0.79
Infertility	23	0.87	0.73, 0.95	63.9 ± 24.0	61.7 ± 25.6	0.41

ICC: intraclass correlation coefficient; EHP-30: Endometriosis Health Profile-30; CI: confidence interval.
[a]Each ICC was significantly different from zero (p < 0.001).
[b]Paired samples t-test, significance two-tailed.

the endometriosis patient group. Participants were recruited from a patient association. Thus, participants with severe forms of endometriosis are likely overrepresented.[34] Recruiting a representative sample of women with endometriosis is a challenge in almost all research settings. Most, if not all, of the EHP-30 validation studies have recruited participants from patient associations and/or from secondary or tertiary referral centers.[10–15] Thus, participants with severe forms of endometriosis are likely overrepresented in all studies, although in varying degree. Moreover, patients attending secondary and tertiary referral centers are more likely to be in active disease and treatment settings, making test–retest reliability analysis difficult. Endometriosis registries would have been a preferable recruitment source to endometriosis associations. However, no endometriosis registry is established in Norway. Furthermore, the responsiveness of the NO-EHP-30 was not evaluated.

The construct self-image does not seem to be measured appropriately by the NO-EHP-30, suggesting a lack of cross-cultural validity of the EHP-30. With multinational and multicultural studies increasing, this study underlines the importance of adequate translation, cross-cultural adaptation, and cross-cultural validation of PROMs.

Declaration of conflicting interests

Funding

References

1. De Graaff AA, D'Hooghe TM, Dunselman GA, et al. The significant effect of endometriosis on physical, mental and social wellbeing: results from an international cross-sectional survey. *Hum Reprod* 2013; 28(10): 2677–2685.

Table 6. Interpretability. Data completeness, mean scores and standard deviations (SD), floor and ceiling effects, and skewness of score distribution.

EHP-30 scales	n	$n_{N/R}$[a]	$n_{missing}$	Mean	SD	Floor effect (%)	Ceiling effect (%)	Coefficient of skewness
Core questionnaire								
Pain	156	N/A	1	34.8	24.2	12.2	0	0.01
Control & powerlessness	156	N/A	1	49.1	26.9	7.7	0.6	−0.36
Emotional well-being	154	N/A	3	40.0	21.5	7.1	0	−0.28
Social support	154	N/A	3	41.3	25.8	13.6	0	−0.18
Self-image	154	N/A	3	39.9	28.8	20.1	1.3	0.06
Modular questionnaire								
Work life	124	30	3	28.3	28.5	33.9	0.8	0.60
Relationship with children	79	74	4	26.4	25.3	34.2	1.3	0.65
Sexual intercourse	136	20	1	48.0	29.3	8.1	3.7	−0.09
Medical profession	88	68	1	38.3	29.9	20.5	2.3	0.15
Treatment	83	73	1	45.0	26.9	10.8	1.2	−0.18
Infertility	61	94	2	62.7	27.1	4.9	9.8	−0.52

N/R: not relevant; N/A: not applicable; EHP-30: Endometriosis Health Profile-30.
[a]Number of participants for whom the scale was not relevant (only applicable for the modular questionnaire).

2. Bottomley A, Pe M, Sloan J, et al. Analysing data from patient-reported outcome and quality of life endpoints for cancer clinical trials: a start in setting international standards. *Lancet Oncol* 2016; 17(11): e510–e514.

3. FDA. Guidance for industry on patient-reported outcome measures: use in medical product development to support labeling claims. *Fed Regist* 2009; 74(235): 65132–65133.

4. Hays RD and Reeve BB. *Measurement and modeling of health-related quality of life*. Amsterdam: Elsevier Inc., 2008.

5. European Medicines Agency (EMA). *Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products*. London: EMA, 2005.

6. McHorney CA, Ware JE Jr and Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993; 31(3): 247–263.

7. Comans TA, Nguyen KH, Mulhern B, et al. Developing a dementia-specific preference–based quality of life measure (AD-5D) in Australia: a valuation study protocol. *BMJ Open* 2018; 8(1): e018996.

8. Jones G, Kennedy S, Barnard A, et al. Development of an endometriosis quality-of-life instrument: the Endometriosis Health Profile-30. *Obstet Gynecol* 2001; 98(2): 258–264.

9. Jones G, Jenkinson C, Taylor N, et al. Measuring quality of life in women with endometriosis: tests of data quality, score reliability, response rate and scaling assumptions of the Endometriosis Health Profile Questionnaire. *Hum Reprod* 2006; 21(10): 2686–2693.

10. van de Burgt TJ, Hendriks JC and Kluivers KB. Quality of life in endometriosis: evaluation of the Dutch-version Endometriosis Health Profile-30 (EHP-30). *Fertil Steril* 2011; 95(5): 1863–1865.

11. Nojomi M, Bijari B, Akhbari R, et al. The assessment of reliability and validity of Persian version of the Endometriosis Health Profile (EHP-30). *Iran J Med Sci* 2011; 36(2): 84–89.

12. Maiorana A, Scafidi Fonti GM, Audino P, et al. The role of EHP-30 as specific instrument to assess the quality of life of Italian women with endometriosis. *Minerva Ginecol* 2012; 64(3): 231–238.

13. Jia S-Z, Leng J-H, Sun P-R, et al. Translation and psychometric evaluation of the simplified Chinese-version Endometriosis Health Profile-30. *Hum Reprod* 2013; 28(3): 691–697.

14. Nogueira-Silva C, Costa P, Martins C, et al. Validation of the Portuguese Version of EHP-30 (The Endometriosis Health Profile-30). *Acta Med Port* 2015; 28(3): 347–356.

15. Chauvet P, Auclair C, Mourgues C, et al. Psychometric properties of the French version of the Endometriosis Health Profile-30, a health-related quality of life instrument. *J Gynecol Obstet Hum Reprod* 2017; 46(3): 235–242.

16. Beaton DE, Bombardier C, Guillemin F, et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000; 25(24): 3186–3191.

17. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010; 19(4): 539–549.

18. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012; 21(4): 651–657.

19. Loge JH, Kaasa S, Hjermstad MJ, et al. Translation and performance of the Norwegian SF-36 Health Survey in patients with rheumatoid arthritis. I. Data quality, scaling assumptions, reliability, and construct validity. *J Clin Epidemiol* 1998; 51(11): 1069–1076.

20. The Language Council of Norway, http://www.sprakradet.no/Vi-og-vart/Om-oss/English-and-other-languages/English/norwegian-bokmal-vs.-nynorsk/ (accessed 1 December 2017).

21. Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the

ISPOR Task Force for Translation and Cultural Adaptation. *Value Health* 2005; 8(2): 94–104.

22. Jenkinson C, Stewart-Brown S, Petersen S, et al. Assessment of the SF-36 version 2 in the United Kingdom. *J Epidemiol Community Health* 1999; 53(1): 46–50.

23. Machin D, Campbell MJ, Tan S-B, et al. *Sample size tables for clinical studies*. 3rd ed. Chichester: John Wiley & Sons, 2009.

24. Mokkink LB, Terwee CB, Patrick DL, et al. 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(7): 737–745.

25. de Vet HC, Ader HJ, Terwee CB, et al. Are factor analytical techniques used appropriately in the validation of health status questionnaires? A systematic review on the quality of factor analysis of the SF-36. *Qual Life Res* 2005; 14(5): 1203–1218; discussion 19–21, 23–24.

26. Ware JE Jr, Kosinski M, Gandek B, et al. The factor structure of the SF-36 Health Survey in 10 countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol* 1998; 51(11): 1159–1165.

27. Reeve BB, Wyrwich KW, Wu AW, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res* 2013; 22(8): 1889–1905.

28. Aaronson N, Alonso J, Burnam A, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002; 11(3): 193–205.

29. Portney LG and Watkins MP. *Foundations of clinical research: applications to practice*. 3rd ed. Harlow: Pearson Education, 2014.

30. Gandek B, Ware JE Jr, Aaronson NK, et al. Tests of data quality, scaling assumptions, and reliability of the SF-36 in eleven countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol* 1998; 51(11): 1149–1158.

31. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; 60(1): 34–42.

32. Fayers PM and Machin D. *Quality of life: the assessment, analysis and interpretation of patient-reported outcomes*. 2nd ed. Chichester: John Wiley & Sons, 2007.

33. The COSMIN checklist with 4-point rating scale, http://www.cosmin.nl/images/upload/files/COSMIN%20checklist%20with%204-point%20scale%2022%20juni%202011.pdf (accessed 1 December 2017).

34. De Graaff AA, Dirksen CD, Simoens S, et al. Quality of life outcomes in women with endometriosis are highly influenced by recruitment strategies. *Hum Reprod* 2015; 30(6): 1331–1341.