Int Urogynecol J (2017) 28:1005–1017
DOI 10.1007/s00192-016-3209-z

ORIGINAL ARTICLE

Norwegian translation, and validation, of the Pelvic Floor Distress Inventory (PFDI-20) and the Pelvic Floor Impact Questionnaire (PFIQ-7)

Catherine J. Teig[1,2,3] · Margreth Grotle[4,5] · Malcolm J. Bond[1] · Cecilia A. C. Prinsen[6] · Marie A. Ellström Engh[2,7] · Milada S. Cvancarova[5,8] · Møyfrid Kjøllesdal[9] · Angelita Martini[1,10]

Received: 9 September 2016 / Accepted: 7 November 2016 / Published online: 6 January 2017

Abstract
Introduction and hypothesis The goal was to translate into Norwegian, and validate, short versions of the Pelvic Floor Distress Inventory (PFDI-20) and the Pelvic Floor Impact Questionnaire (PFIQ-7) using a sample of women with symptomatic pelvic organ prolapse and pelvic floor dysfunction.
Methods Modified European Organization for Research and Treatment of Cancer Guidelines were used for translation and cultural adaptation. Of 212 eligible Norwegian women who consented to participate, 205 completed the questionnaires, of whom 50 were retested after 1 – 3 weeks, and 76 were tested 6 months after surgery. Reliability, validity and responsiveness were evaluated. Additionally, interpretability, the smallest detectable change, the standard error of measurement, floor and ceiling effects, and the percentages of missing items are reported.
Results Reliability ranged from 0.66 to 0.93 and intraclass correlation coefficients from 0.85 to 0.94. Both construct validity and responsiveness were found to be adequate. The responsiveness of the PFDI-20 was further supported by areas under the curve above 0.70. Estimates were lower for the PFIQ-7. The smallest detectable changes at the individual level were 15 – 21 % and 17 – 27 % for the PFDI-20 and PFIQ-7, respectively. The absolute values of the minimal important changes in the total scores were 48 and 47, respectively. No floor or ceiling effects were evident in the distributions of the PFDI-20 and PFIQ-7 total scores.
Conclusions The translated questionnaires provided adequate reliability, validity and good responsiveness to change. These short versions of the PFDI and PFIQ are robust measuring instruments that will enable symptom severity and health-related quality of life to be evaluated in the Norwegian context.

Keywords Pelvic organ prolapse · Pelvic floor dysfunction · PFDI-20 · PFIQ-7 · Quality of life · Questionnaire translation

✉ Catherine J. Teig
chpl@ahus.no

[1] School of Medicine, Flinders University, Adelaide, Australia

[2] Department of Obstetrics and Gynaecology, Akershus University Hospital, Lørenskog, Norway

[3] The Pelvic Floor Centre, Division of Surgery, Akershus University Hospital, 1000, 1478 Lørenskog, Norway

[4] Oslo and Akershus University College of Applied Sciences, Faculty of Health Sciences, Department of Physiotherapy, Oslo University Hospital, Oslo, Norway

[5] FORMI, Clinic for Surgery and Neurology, Oslo University Hospital, Oslo, Norway

[6] VU University Medical Center, Department of Epidemiology and Biostatistics, Institute for Health and Care Research, Amsterdam, The Netherlands

[7] Faculty Division Akershus University Hospital, University of Oslo, Oslo, Norway

[8] Oslo and Akershus University College of Applied Sciences, Faculty of Health Sciences, Department of Nursing and Health Promotion, Oslo University Hospital, Oslo, Norway

[9] Drammens Gynaecology Private Clinic, Drammen, Norway

[10] Centre for Health Services Research, School of Population Health, University of Western Australia, Perth, Australia

Introduction

Pelvic organ prolapse (POP) is defined as the symptomatic descent of one or more of the anterior vaginal wall, the posterior vaginal wall, and the apex of the vagina (uterus or vault) or uterus [1]. Other pelvic floor dysfunctions (PFD) often coexist with POP, such as lower urinary tract, bowel and sexual dysfunctions. POP and other PFD affect a substantial proportion of women [2] and can often cause bothersome symptoms and have a negative effect on psychological and social wellbeing [3]. To better understand a patient's condition and the effect of treatment, patient-reported outcomes such as condition-specific health-related quality of life (HRQoL) are often assessed [3].

Two common instruments used for this purpose are the Pelvic Floor Distress Inventory and the Pelvic Floor Impact Questionnaire [3], for which abbreviated 20-item (PFDI-20) and 7-item (PFIQ-7) versions, respectively, have been validated to reduce the burden on participants [4]. Both have been tested for reliability, validity and responsiveness to change against their original longer counterparts, and demonstrated moderate to excellent associations [4, 5]. The PFDI-20 assesses the presence of symptoms and bother in three domains (POP, bowel and urinary), and the PFIQ-7 assesses the impact on HRQoL in these domains. Both the PFDI-20 and PFIQ-7 are designed to evaluate the efficacy of therapy and have been shown to discriminate between women with and without improvement following treatment [4, 5].

The PFDI-20 and PFIQ-7 are highly recommended (grade A) [6] and although validated in several languages [7–9], there are as yet no Norwegian versions. Therefore, the aims of the current study were to translate the PFDI-20 and PFIQ-7 into Norwegian and test their measurement properties (reliability, validity and responsiveness to change) in a prospective longitudinal study of women with POP and PFD in the tertiary setting.

Materials and methods

Ethics approval

Approval was granted by the regional committees for Medical and Health Research Ethics (Norway) and the Flinders University Social and Behavioural Research Ethics Committee (Australia). Permission was also granted by the developer of both instruments. Written informed consent was obtained from all participants.

Translation and cultural adaptation

The PFDI-20 and PFIQ-7 were first translated from English into Norwegian using a multistep translation and cultural adaptation method. This new method combined the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Group Guidelines [10], the Delphi method [11, 12] and expert panel review [13]. It involved two independent forward and back translations [10], with the addition of the Delphi method [11, 12] (i.e. anonymous voting, controlled feedback, and statistical group response), to establish consensus on the translated items among a panel of bilingual pelvic floor experts comprising gynaecologists, colorectal surgeons, a urologist, a physiotherapist and a urotherapist [13]. The translated instruments were then pilot tested for comprehensibility, readability and equivalence through face-to-face semistructured interviews with 20 women with POP (with or without urinary or bowel dysfunction). Minor discrepancies were identified and amended, resulting in comprehensible Norwegian versions of the PFDI-20 and PFIQ-7 with readability level at a reading age of 12 years. These are included in the Appendices 1 and 2.

Participants and procedure

Participants were patients recruited through the Department of Obstetrics and Gynaecology of Akershus University Hospital, Norway, from June 2014 to September 2015. Two cohorts were included: those with POP (nonsurgical patients), and those undergoing surgery for POP (surgical patients; Table 1). For inclusion nonsurgical patients had to be referred to the Outpatient Department with symptomatic POP (with or without urinary or bowel dysfunction), while the surgical patients had in addition to have anatomic POP Quantification (POP-Q) [14] stage 2 – 4 and to be scheduled for vaginal repair.

Exclusion criteria were age less than 18 years, inability to understand Norwegian and/or complete a patient-reported outcome questionnaire, and visual impairment. The sample size was based on the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) recommendations of a minimum of 50 participants for every subgroup analysis except internal consistency [15], which was based on a subject-to-item ratio of at least 4:1 (minimum 108 participants) [15].

The participants completed the PFDI-20, PFIQ-7 and the SF-36v2 Norwegian Health Survey (SF36) [16] at baseline (T0), and a subsample completed the questionnaires 1 – 3 weeks later (T1). This interval was chosen on the assumption that it would be short enough for the participants' POP condition to remain unchanged, but long enough to ensure that they would not recall their T0 responses. Patients scheduled for POP surgery also completed the questionnaires

Table 1 Baseline characteristics of the participants and summary statistics for key study variables

	Total sample ($N = 205$)	Nonsurgical participants ($N = 109$)	Surgical participants ($N = 96$)
Body mass index (kg/m^2), median (range)	26 (17 – 45)	25 (17 – 45)	27 (18 – 45)
Parity, median (range)	2 (1 – 8)	2 (1 – 8)	2 (1 – 7)
POP-Q stage, *n* (%)[a]			
1	7 (3.4)	3 (3.1)	4 (3.7)
2	83 (40.5)	40 (41.7)	43 (39.4)
3	77 (37.6)	35 (36.5)	42 (38.5)
4	8 (3.9)	3 (3.1)	5 (4.6)
Category of POP stage 2 – 4, *n* (%)			
Cystocele (anterior compartment)	154 (75.1)	76 (79.2)	78 (71.6)
Rectocele (posterior compartment)	111 (54.1)	53 (55.2)	58 (53.2)
Apical prolapse (middle compartment)	32 (15.6)	13 (13.5)	19 (17.4)
Previous pelvic reconstructive surgery, *n* (%)	41 (20.0)	19 (19.8)	22 (20.2)
Previous hysterectomy, *n* (%)	43 (21.0)	21 (21.9)	22 (20.2)
Surgical procedure ($n = 76$), *n* (%)[b]			
Anterior colporrhaphy	42 (55.3)		42 (55.3)
Posterior colporrhaphy	22 (28.9)		22 (28.9)
Manchester operation	3 (3.9)		3 (3.9)
Vaginal hysterectomy	17 (22.4)		17 (22.4)
Sacrospinous fixation	2 (2.6)		2 (2.6)
Isolated amputation of the cervix	4 (5.3)		4 (5.3)
Enterocele operation	1 (1.3)		1 (1.3)
Questionnaire scores, mean (SD)			
PFDI-20	107.7 (54.3)	106.3 (55.2)	108.9 (54.2)
POPDI-6	45.5 (22.3)	47.7 (23.5)	43.5 (22.4)
CRADI-8	24.9 (18.9)	25.7 (20.1)	24.2 (19.8)
UDI-6	37.4 (24.6)	36.1 (25.3)	38.6 (24.9)
PFIQ-7	60.9 (53.8)	59.2 (55.2)	62.4 (54.7)
UIQ-7	25.4 (23.2)	26.2 (24.1)	24.7 (23.9)
CRAIQ-7	14.2 (19.8)	12.1 (20.3)	16.1 (20.1)
POPIQ-7	21.2 (22.4)	21.6 (23.1)	20.8 (22.9)

[a] Several patients had POP in more than one compartment. The highest stage reported in any compartment is recorded

[b] Several patients underwent more than one surgical procedure

6 months after surgery (T2). At T0 participants provided sociodemographic data (age, gender, parity), body mass index and previous surgery data as sample descriptors. A POP-Q examination was performed at both T0 and T2. Figure 1 shows a flow chart of patient recruitment and participation.

Measurement instruments

The 20-item PFDI-20 measures symptom distress during the past 3 months. Responses are given on a scale ranging from 0 ('no') to 4 ('yes, quite a bit') [5]. Three subscales are also available: the Urinary Distress Inventory (UDI-6), the Pelvic Organ Prolapse Distress Inventory (POPDI-6), and the Colorectal–Anal Distress Inventory (CRADI-8). The total score is converted to a range of 0 to 300, and the subscales are scored 0 to 100. In all cases higher scores indicate greater distress. The seven-item PFIQ-7 measures HRQoL issues in women with PFD (including daily physical/social activity, travel, and emotional health) during the past 3 months. Responses are given on a scale ranging from 0 ('not at all')

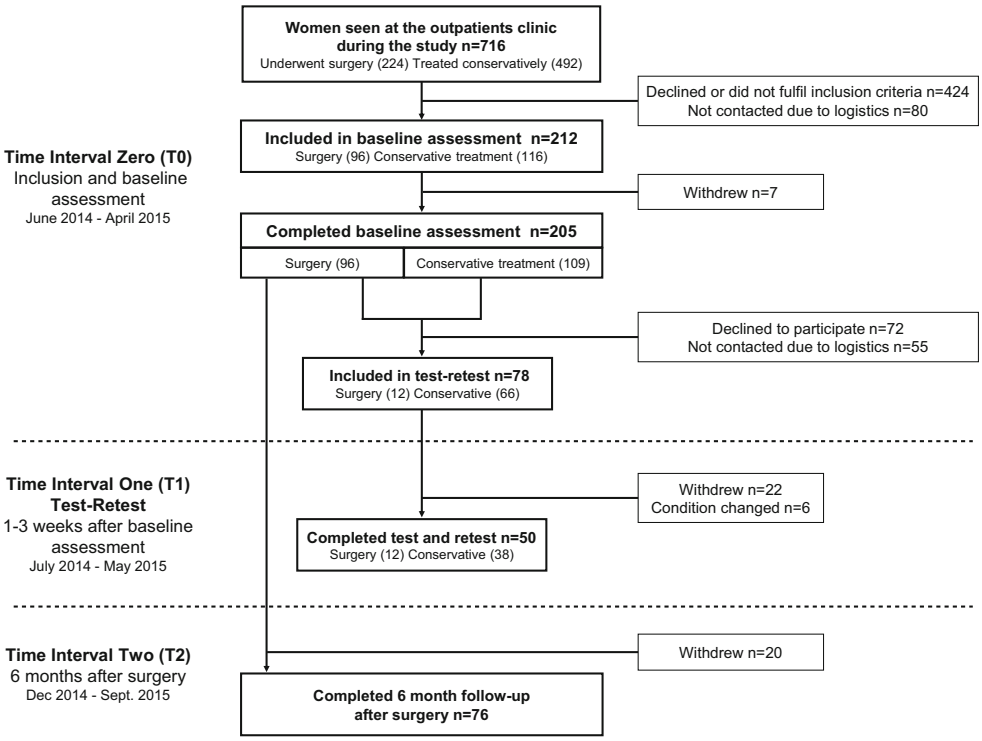

Fig. 1 Flow chart of patient recruitment and participation

to 3 ('quite a bit'). The PFIQ-7 also has three subscales: the Urinary Impact Questionnaire (UIQ-7), the Pelvic Organ Prolapse Impact Questionnaire (POPIQ-7), and the Colorectal–Anal Impact Questionnaire (CRAIQ-7) [5]. Again, the total score is converted to a range of 0 to 300, and the subscales are scored 0 – 100. Higher scores indicate greater symptom distress and impact on the patient's HRQoL [5]. The SF36 is a multipurpose generic health outcome measure comprising 36 items. For the current study, only the Physical Component Summary (PCS) score and the Mental Health Component Summary (MCS) score are reported. For both PCS and MCS, lower scores indicate poorer health [17].

At retest (T1), participants were asked if their condition had changed during the interim period [18] with the question 'Compared to the first time you completed the questionnaires, has your prolapse condition changed?' (if 'Yes', women were excluded from the retest). At T2, participants were also asked 'In general, how much did the treatment improve your pelvic organ prolapse?'(global rating of change, GRC). Responses are given on a six-point scale from 'improved significantly' to 'no significant improvement' [18].

Statistical methods

Analyses were conducted using SPSS version 22.0 (IBM Corp., Armonk, NY). Statistical significance was assumed at $p < 0.05$. COSMIN recommendations were used as a guide for evaluating the measurement properties of the Norwegian PFDI-20 and PFIQ-7 [19, 20]. First, floor and ceiling effects were examined and considered problematic if more than 15 % of participants achieved the highest or lowest possible score [15]. Missing data at the item level were also noted. Based on COSMIN recommendations, <3 % is acceptable and >15 % is unacceptable [18]. Cronbach's alpha was calculated for PFDI-20 and PFIQ-7 scores as a measure of internal consistency (the degree of interrelatedness among the items [20]). A value of 0.70 or greater is considered to indicate adequate internal consistency [15, 21].

Test–retest reliability (the degree to which a measurement is free from error [20]) was evaluated using intraclass correlation coefficients (ICCs) to quantify the agreement between PFDI-20 and PFIQ-7 scores [15, 19]. ICCs were calculated according to the method of McGraw and Wong [15]. Coefficients of at least 0.70 are considered adequate [15, 18]. Measurement error (the systematic and random error of a patient's score that cannot be attributed to true changes in the construct being measured) [20] was also assessed. It is considered acceptable when the smallest detectable change (SDC; $1.96 \times \sqrt{2} \times$ SEM, where SEM is the standard error of measurement) is smaller than the minimal important change (MIC) [15]. SEM was calculated as the square root of the variance from analysis of variance, including systematic differences (SEM agreement) [15].

The degree to which the scores of a measurement instrument are consistent with hypotheses based on the assumption that the measurement instrument validly measures the

Table 2 Confirmation or rejection of baseline hypotheses

Hypothesis tested		Correlation coefficient (r)	Confirmed?
Correlation expected	Between		
High positive[a]	1. PFDI-20 and PFIQ-7 total scores	0.75	Yes
	2. PFDI-20 POPDI-6 and PFIQ-7 POPIQ-7	0.58	No
	3. PFDI-20 CRADI-8 and PFIQ-7 CRADIQ-7	0.68	Yes
	4. PFDI-20 UDI-6 and PFIQ-7 UIQ-7	0.76	Yes
Moderate negative[b]	5. PFDI-20 total score and SF36 PCS	−0.33	Yes
	6. PFIQ-7 total score and SF36 MCS	−0.33	Yes
	7. PFIQ-7 total score and SF36 PCS	−0.44	Yes
Low[c]	8. PFDI-20 total score and SF36 MCS	0.22	Yes

MCS Mental health component summary score, *PCS* Physical component summary score

[a] The PFDI-20 and PFIQ-7 measure the same construct [7]

[b] The PFDI-20 and PFIQ-7 subscales and the SF36 MCS/PCS components appear to measure similar but not equivalent constructs [7]:

[c] The PFDI-20 and SF36 MCS do not appear to measure similar constructs [7]:

construct to be measured [20] was assessed by testing eight hypotheses expressed in terms of the expected direction and magnitude of the effect (Table 2). Correlations were calculated between the PFDI-20 and PFIQ-7 scores and the SF36 at baseline [15]. Both convergent and divergent validity were tested [18], with the expectation that correlations between related constructs would be high, while those between unrelated constructs would be low or non-existent [18]. Coefficients were arbitrarily considered low (<0.30), moderate (0.30 – 0.59) or high (≥0.60).

Responsiveness to change (the ability to detect change over time in the construct being measured [20]) of the PFDI-20 and PFIQ-7 was assessed by addressing five hypotheses (Table 3), tested by correlating changes in PFDI-20 and PFIQ-7 scores with changes in SF36 scores [18]. Each questionnaire was considered responsive if at least 75 % of the relevant hypotheses were supported [15]. It was expected that correlations among related constructs would be higher than among unrelated constructs [18]. Compared with the PFDI-20 and PFIQ-7, the SF36 should be relatively unresponsive to change in women undergoing POP surgery [5]. Further, receiver operating characteristic (ROC) curves were constructed and the areas under the curves (AUC) calculated [18]. Changes in scores between T0 and T2 were calculated. After surgery, patients who reported being 'much improved' or 'greatly improved' in their responses to the GRC [22, 23] were classified as 'improved significantly' while those who reported 'little improvement' or 'no change' were classified as 'no significant improvement' [18] (Table 4). Women who reported deterioration in the GRC were excluded from the responsiveness analyses. The PFDI-20 and PFIQ-7 were considered to be responsive to change if AUCs exceeded 0.70 [18].

The MIC, a measure of the interpretability of the change in score, was also calculated [18]. It was determined by the anchor-based MIC distribution, using the ROC approach [18]. The optimal ROC cut-off points were taken as the value

Table 3 Confirmation or rejection of responsiveness hypotheses

Hypothesis tested		Correlation coefficient (r)	Confirmed?
Correlation expected	Between		
High[a]	1. PFDI-20 and PFIQ-7 total change scores	0.65	Yes
Moderate negative[b]	2. PFDI-20 total and SF36 PCS change scores	−0.42	Yes
	3. PFIQ-7 total and SF36 PCS change scores	−0.34	Yes
Low[c]	4. PFDI-20 total and SF36 MCS change scores	0.15	Yes
	5. PFIQ-7 total and SF36 MCS change scores	0.14	Yes

MCS Mental health component summary score, *PCS* Physical component summary score

[a] The PFDI-20 and PFIQ-7 measure the same construct [4, 5]

[b] The PFDI-20 and PFIQ-7 appear to measure similar but not equal constructs to the SF36 PCS component [4, 5]

[c] The PFDI-20 and SF36 MCS do not appear to measure similar constructs [4, 5]

Table 4 Responsiveness and interpretability of the PFDI-20 and PFIQ-7 in terms of the changes in total scores from T0 to T1 in 76 women completing the 6-month follow-up (T2)

Global rating of change	Number (%) of women	Change in score, mean (SD)[a]	
		PFDI-20	PFIQ-7
Improved significantly	66 (89)	−63 (44.2)	−49 (50.5)
No significant improvement	8 (11)	−0.4 (66.7)	−36 (53.1)
Missing cases	2 (0.3)	–	–
AUC (95 % confidence interval)		0.74 (0.600 – 0.928)	0.586 (0.345 – 0.826)
p value		0.035	0.459
MIC		−48	−47
Sensitivity/specificity for MIC estimate		0.839/0.701	0.763/0.672

MIC Minimal important change

[a] PFDI-20 and PFIQ-7 total scores range from 0 to 300. Negative scores indicate a reduction in distress and/or impact of symptoms

for which the sum of the proportions of misclassification, i.e. (1 − sensitivity) + (1 − specificity), was smallest [9]. The MIC must be bigger than the SDC for a change in score to be distinguishable from measurement error. Interpretation of change scores was tested using the anchor-based MIC distribution method to assess which changes from PFDI-20 and PFIQ-7 total scores correspond with the MIC defined on the anchor (i.e. GRC), which distinguished patients who had 'improved significantly' after surgery from those who showed 'no significant improvement [18].

Results

During the study period 716 consecutive patients were referred to the outpatient clinic for POP. Of these, 424 (58 %) did not meet the inclusion criteria or declined to participate. A further 80 (13 %) were not invited to participate for logistical reasons (Fig. 1), leaving 212 eligible women (29 %) who consented to participate. Of these, 205 completed the questionnaires at T0 giving an excellent response rate of 96.7 %. A subsample of 56 women (27.3 %) completed questionnaires at T1. Of the 96 women undergoing surgery, 76 (79.1 %) completed the questionnaires at T2. The retest evaluation (T1) was completed a median of 11 days (range 6 – 21 days) after T0. At T1 six patients indicated a change in the symptoms and severity of their POP and were not considered further in the study (Fig. 1). The T2 evaluation was completed a median of 184 days (range 153 – 189 days) after T0.

The median age of the women was 61 years (range 27 – 82 years). The majority of women with POP had POP-Q stage 2 or 3. Anterior compartment prolapse was the most common type of POP. Several women had POP in more than one compartment. Women who were treated surgically underwent only vaginal repair. Anterior and posterior compartment repair were the most common procedures (Table 1). Of the 205 women, 172 (83.9 %) completing the PFDI-20 reported symptoms in all three PFD domains, 27 (13.2 %) reported symptoms in two PFD domains, and 6 (2.9 %) reported symptoms in only one domain. All 205 women completing the PFDI-20 reported symptoms of POP[1], 192 women (94 %) reported lower urinary tract symptoms[2] and 184 women (88 %) reported bowel symptoms[3].

Evaluation of measurement properties

No floor or ceiling effects were found in the distributions of the PFDI-20 and PFIQ-7 total scores (Table 5). Similarly, no ceiling effect was observed for any of the PFDI-20 or PFIQ-7 subscales. However, the UIQ-7 subscales showed small floor effects, while major floor effects were noted for the POPIQ-7 and CRAIQ-7 subscales.

Missing data at baseline were associated with only 0.82 % of PFDI-20 items and 1.92 % of PFIQ-7 items. Cronbach's alpha for the PFDI-20 and PFIQ-7 total scores was 0.83 and 0.93, respectively, demonstrating very satisfactory internal consistency. Similarly, subscale coefficients (Table 6) were generally satisfactory to excellent, with the exception of POPDI-6 (0.66). In all cases, for both scales, test–retest ICCs (Table 6) indicated adequate reliability ($p < 0.001$ for all coefficients). The SDC at the individual level was 16.7 (16.7 %) to 26.3 (26.3 %) for the PFDI-20 subscales (range 0 – 100), and was 46.1 for the PFDI-20 total score (range 0 – 300), i.e. a relative SDC of 15.3 %

[1] Based on a sensation of a bulge in the pelvic area (i.e. PFDI-20)

[2] Based on lower urinary tract symptoms (i.e. PFDI-20)

[3] Based on bowel symptoms (i.e. PFDI-20) .

Table 5 Floor and ceiling effects of baseline scores

Measurement instrument	Score range	Floor, *n* (%)	Ceiling, *n* (%)
PFDI-20	0 – 300	0 (0)	0 (0)
POPDI-6	0 – 100	0 (0)	14 (7)
CRADI-8	0 – 100	1 (0.5)	6 (3)
UDI-6	0 – 100	0 (0)	14 (7)
PFIQ-7	0 – 300	14 (7)	0 (0)
POPIQ-7	0 – 100	52 (26)	0 (0)
CRAIQ-7	0 – 100	94 (47)	0 (0)
UIQ-7	0 – 100	39 (19.5)	0 (0)

of the total score. For the PFIQ-7, the SDCs were slightly larger. The SDC was 26.1 (26.1 %) to 27.2 (27.2 %) for the PFIQ-7 subscales (range 0 – 100), and was 62.1 for the PFIQ-7 total score (range 0 – 300), i.e. a relative SDC of 20.7 % of the total score (Table 6).

Construct validity was adequate, with 88 % of predefined hypotheses (seven of eight) confirmed (Table 2). The exception was the association between POPDI and POPIQ-7, with only a moderate positive correlation (0.58). In all other cases, as hypothesized, measures of the same construct provided high positive correlations. Further, scales measuring similar, but not equivalent, constructs showed moderate correlations, and scales measuring unrelated constructs showed low correlations (Table 2).

Responsiveness was adequate, with 100 % of the predefined hypotheses (five of five) confirmed (Table 3). Change in scores measuring the same construct showed high positive correlations, those measuring similar but not equivalent constructs showed moderate negative correlations, and those measuring unrelated constructs showed low correlations. Responsiveness to changes in PFDI-20 scores was further supported by AUC values of ≥0.70, whereas the AUCs were lower for changes in PFIQ-7 scores (Table 4). The MIC for the PFDI-20 total score (0 – 300) was 48, which was slightly larger than the SDC (46.01; Table 6). This suggests that an improvement in PFDI-20 score of ≥48 can be regarded as a clinically relevant change. Patients who had 'improved significantly' on the GRC 6 months after surgery achieved a mean change of 63, indicating clinically relevant improvement. The absolute value of MIC for the PFIQ-7 total score (0 – 300) was 47, which was smaller than the SDC (62.1; Table 6). Hence, a score of ≤47 points cannot be considered a clinically relevant improvement. While such a change may be considered important by the patient, it cannot be distinguished from measurement error.

Discussion

Norwegian translations of the PFDI-20 and PFIQ-7 were found to have adequate reliability (test/retest reliability, and internal consistency), validity and responsiveness to change in a homogeneous sample of women at baseline and after surgical treatment. As predicted [5], all retest assessments of the PFDI-20 and PFIQ-7 showed adequate reliability. In general, internal consistency was at least adequate, with the exception of the POPDI-6, for which internal consistency was found to be less than adequate (0.66). Interestingly, some cross-cultural adapted versions have shown a similar issue for the POPDI-6 [7, 9].

As in Swedish and Dutch studies [7, 9], no ceiling effects were found for total or subscale scores of these measures. However, as floor effects were found in the PFIQ-7 POPIQ and CRAIQ-7, it is suggested that the PFIQ-7

Table 6 Internal consistency and test–retest statistics

Measurement instrument	Cronbach's alpha	Reliability		Change in score		
		Intraclass correlation coefficient	95 % confidence interval	Mean (SD)	Standard error of measurement (%)	Smallest detectable change (%)
PFDI-20	0.83	0.944	0.897 – 0.969	12.3 (23.5)	16.7 (5.6)	46.1 (15.3)
POPDI-6	0.66	0.895	0.807 – 0.943	4.2 (13.4)	9.5 (9.5)	26.3 (26.3)
CRADI-8	0.72	0.938	0.887 – 0.966	6.0 (8.5)	6.0 (6.0)	16.7 (16.7)
UDI-6	0.71	0.918	0.849 – 0.955	2.2 (11.5)	8.1 (8.1)	22.5 (22.5)
PFIQ-7	0.93	0.899	0.821 – 0.943	13.0 (31.7)	22.4 (7.5)	62.1 (20.7)
POPIQ-7	0.88	0.891	0.807 – 0.938	4.8 (13.6)	9.6 (9.6)	26.7 (26.7)
CRAIQ-7	0.91	0.852	0.737 – 0.916	3.8 (13.9)	9.8 (9.8)	27.2 (27.2)
UIQ-7	0.88	0.903	0.827 – 0.945	4.5 (13.3)	9.4 (9.4)	26.1 (26.1)

should be interpreted in terms of both the total score and subscale scores. This supports the findings of the Dutch study, which found similar floor effects [9]. The authors pointed out that patients can experience various types of PFDs, but might not experience all associated symptoms (e.g. POP and defecation problems without urinary incontinence) [9].

Responsiveness was high for PFDI-20 and moderate for PFIQ-7. Thus, the PFDI-20 exhibited a better ability to capture change. For the ROC curve analysis, the patients were divided into two groups: 'no significant improvement' and 'improved significantly'. During sensitivity analysis using the ROC method, two patients who reported 'no change' were included in the combined 'no significant improvement' category. Further, we redefined minimal importance and dichotomized GRC as 'improved slightly'/'much improved and improved greatly' [24]. The dichotomization into the two GRC categories resulted in similar responsiveness for the PFDI-20 and PFIQ-7. Moreover, the results for the PFDI-20 were similar to those in a Danish translation study, which also showed that the instrument has adequate responsiveness to change [24].

GRC might be seen as not measuring the same constructs as the PFDI-20 and the PFIQ-7 scales. However, Gelhorn et al. [22] consider that the PFDI-20, PFIQ-7 and GRC (which they refer to as Patient Global Impression of Change) are sound external measures of patients' perception of change. The PFDI-20 showed a MIC of 48, which is similar to the minimally clinically importance difference of 45 points found by Barber et al. [5]. The PFDI-20 can detect clinically relevant improvement, whereas the measurement error of PFIQ-7 was too large to detect clinically relevant improvement. The Dutch studies found similar results for both the PFDI-20 and PFIQ-7 [9].

Some caveats to the interpretation of the current results should be acknowledged. First, a limitation was the recruitment of only those women with symptomatic POP (with or without urinary or bowel dysfunction). That is, women with only urinary or bowel dysfunction were not recruited. However, both urinary and bowel dysfunction were present with high frequency in the total sample, with only six participants (2.9 %) reporting having exclusively POP. In terms of psychometrics, validation data were collected only within a tertiary setting, which limits generalizability. Further validation studies in more general contexts are therefore recommended. Further recommendations include responsiveness testing for conservative treatment, and establishing confirmatory factor analysis and clinically meaningful interpretations of PFDI-20 and PFIQ-7 total scores and subscales. Educational level was not included in the baseline characteristics and the study was not able to demonstrate if the questionnaires could be understood by women of all educational levels. Moreover, during the pilot test sexuality was an aspect identified as important to patients and not covered in the PFDI-20 and PFIQ-7. Employing a third measuring instrument covering sexuality issues for women with PFD should also be considered [25]. Finally, validation of electronic administration versions of the PFDI-20 and PFIQ-7 is also recommended in clinical and research settings [26]. Electronic administration may encourage higher survey response rates and, hence, reduce nonresponse bias.

Conclusions

The translated and validated Norwegian versions of the PFDI-20 and PFIQ-7 are effective measures of symptom distress and quality of life among Norwegian women with POP and PFD. The PFDI-20 exhibited a better ability to capture changes than the PFIQ-7. The use of these instruments in the clinical and research settings will provide data that could lead to better patient management and policy decisions in Norway.

Acknowledgements The authors acknowledge the patients and expert panel who provided input to this study; the Department of Obstetrics and Gynaecology, Akershus University Hospital, for patient recruitment; A.S. Midthaug and K.A. Askestad for assistance with data collection; M. Solberg for assistance with data entry; Dr. W. Mokkink for assistance with editing; Drs. T. Øresland, K. Bø, H.C.W. de Vet and S. Saga for assistance with protocol development; and M.D. Barber for permission to translate the PFDI-20 and PFIQ-7 and assistance in the translation process.

Compliance with ethical standards

Funding This study was funded by the Division of Surgery, Akershus University Hospital, and the Strategic Research Foundation (*Strategiske Forskningsmidler*), Akershus University Hospital. The Strategic Research Foundation was not involved in the design, conduct, analysis or interpretation of the study, or the review or approval of the manuscript.

Conflicts of interest None.

Appendix 1: Spørreskjema om bekkenbunnsplager - skjema PFDI-20

Veiledning: Vennligst svar på alle spørsmålene i spørreskjemaet. Spørsmålene dreier seg om hvorvidt du har visse symptomer i tarmen, blæren eller bekkenregionen, og i så fall hvor mye de plager deg. Svar på spørsmålene ved å krysse av i den eller de boksene som passer for deg. Hvis du er usikker på hva du skal svare, svarer du så godt du kan. Vær snill og svar på spørsmålene ut fra de symptomer du har hatt gjennom de siste tre månedene.

1. Kjenner du ofte *trykk* i nedre del av magen?

☐ Nei; ☐ Ja
0

<u>Hvis ja</u>, hvor mye plager det deg?

☐ **1**	☐ **2**	☐ **3**	☐ **4**
Ikke i det hele tatt -	Litt -	I noen grad -	Ganske mye

2. Har du ofte *tyngdefølelse i bekkenet*?

☐ Nei; ☐ Ja
0

<u>Hvis ja</u>, hvor mye plager det deg?

☐ **1**	☐ **2**	☐ **3**	☐ **4**
Ikke i det hele tatt -	Litt -	I noen grad -	Ganske mye

3. Kjenner eller ser du ofte noe som buler eller faller ut i skjeden?

☐ Nei; ☐ Ja
0

<u>Hvis ja</u>, hvor mye plager det deg?

☐ **1**	☐ **2**	☐ **3**	☐ **4**
Ikke i det hele tatt -	Litt -	I noen grad -	Ganske mye

4. Må du ofte presse med fingre i skjeden eller rundt endetarmsåpningen for å få ut avføring eller få tømt tarmen helt?

☐ Nei; ☐ Ja
0

<u>Hvis ja</u>, hvor mye plager det deg?

☐ **1**	☐ **2**	☐ **3**	☐ **4**
Ikke i det hele tatt -	Litt -	I noen grad -	Ganske mye

5. Føler du ofte at du ikke får tømt blæren helt?

☐ Nei; ☐ Ja
0

<u>Hvis ja</u>, hvor mye plager det deg?

☐ **1**	☐ **2**	☐ **3**	☐ **4**
Ikke i det hele tatt -	Litt -	I noen grad -	Ganske mye

6. Hender det at du må trykke inn med fingrene noe som buler i skjeden, for å få tisset eller tømt blæren helt?

☐ Nei; ☐ Ja
0

<u>Hvis ja</u>, hvor mye plager det deg?

☐ **1**	☐ **2**	☐ **3**	☐ **4**
Ikke i det hele tatt -	Litt -	I noen grad -	Ganske mye

7. Føler du at du må presse for hardt for å få ut avføringen?

☐ Nei; ☐ Ja
0

Hvis ja, hvor mye plager det deg?

☐ 1	☐ 2	☐ 3	☐ 4
Ikke i det hele tatt	Litt	I noen grad	Ganske mye

8. Føler du at du ikke har tømt tarmen helt, når du har hatt avføring?

☐ Nei; ☐ Ja
0

Hvis ja, hvor mye plager det deg?

☐ 1	☐ 2	☐ 3	☐ 4
Ikke i det hele tatt	Litt	I noen grad	Ganske mye

9. Har du ofte avføringslekkasje når avføringen er fast?

☐ Nei; ☐ Ja
0

Hvis ja, hvor mye plager det deg?

☐ 1	☐ 2	☐ 3	☐ 4
Ikke i det hele tatt	Litt	I noen grad	Ganske mye

10. Har du ofte avføringslekkasje når avføringen er løs eller flytende?

☐ Nei; ☐ Ja
0

Hvis ja, hvor mye plager det deg?

☐ 1	☐ 2	☐ 3	☐ 4
Ikke i det hele tatt	Litt	I noen grad	Ganske mye

11. Har du ofte ufrivillig lekkasje av luft fra tarmen?

☐ Nei; ☐ Ja
0

Hvis ja, hvor mye plager det deg?

☐ 1	☐ 2	☐ 3	☐ 4
Ikke i det hele tatt	Litt	I noen grad	Ganske mye

12. Har du ofte smerter når du har avføring?

☐ Nei; ☐ Ja
0

Hvis ja, hvor mye plager det deg?

☐ 1	☐ 2	☐ 3	☐ 4
Ikke i det hele tatt	Litt	I noen grad	Ganske mye

13. Opplever du så sterk avføringstrang at du må løpe til toalettet?

☐ Nei; ☐ Ja
0

Hvis ja, hvor mye plager det deg?

☐ 1	☐ 2	☐ 3	☐ 4
Ikke i det hele tatt	Litt	I noen grad	Ganske mye

14. Hender det at en del av tarmen følger med ut gjennom endetarmsåpningen under eller etter avføring?

☐ Nei; ☐ Ja
0

<u>Hvis ja</u>, hvor mye plager det deg?

☐ **1**	☐ **2**	☐ **3**	☐ **4**
Ikke i det hele tatt -	Litt -	I noen grad -	Ganske mye

15. Har du vanligvis hyppig vannlating?

☐ Nei; ☐ Ja
0

<u>Hvis ja</u>, hvor mye plager det deg?

☐ **1**	☐ **2**	☐ **3**	☐ **4**
Ikke i det hele tatt -	Litt -	I noen grad -	Ganske mye

16. Opplever du så sterk vannlatingstrang at du ikke rekker til toalettet før du får lekkasje?

☐ Nei; ☐ Ja
0

<u>Hvis ja</u>, hvor mye plager det deg?

☐ **1**	☐ **2**	☐ **3**	☐ **4**
Ikke i det hele tatt -	Litt -	I noen grad -	Ganske mye

17. Har du ofte urinlekkasje når du hoster, nyser eller ler?

☐ Nei; ☐ Ja
0

<u>Hvis ja</u>, hvor mye plager det deg?

☐ **1**	☐ **2**	☐ **3**	☐ **4**
Ikke i det hele tatt -	Litt -	I noen grad -	Ganske mye

18. Har du ofte små urinlekkasjer (dvs. dråper)?

☐ Nei; ☐ Ja
0

<u>Hvis ja</u>, hvor mye plager det deg?

☐ **1**	☐ **2**	☐ **3**	☐ **4**
Ikke i det hele tatt -	Litt -	I noen grad -	Ganske mye

19. Har du ofte problemer med å tømme blæren?

☐ Nei; ☐ Ja
0

<u>Hvis ja</u>, hvor mye plager det deg?

☐ **1**	☐ **2**	☐ **3**	☐ **4**
Ikke i det hele tatt -	Litt -	I noen grad -	Ganske mye

20. Har du ofte *smerte* eller *ubehag* i nedre del av magen eller underlivet?

☐ Nei; ☐ Ja
0

<u>Hvis ja</u>, hvor mye plager det deg?

☐ **1**	☐ **2**	☐ **3**	☐ **4**
Ikke i det hele tatt -	Litt -	I noen grad -	Ganske mye

Appendix 2: Spørreskjema om bekkenbunnsplager og innvirkning på dagliglivet - skjema PFIQ-7

Veiledning: Noen kvinner opplever at symptomer fra blæren, endetarmen eller skjeden påvirker deres gjøremål, forhold og følelser. For hvert av spørsmålene ber vi deg krysse av for svaret som best beskriver hvordan dine gjøremål, forhold eller følelser har blitt påvirket av symptomer eller plager fra blære, endetarm eller skjede de tre siste månedene. Husk å krysse av i alle de tre kolonnene for hvert spørsmål.

Hvordan pleier symptomer eller plager fra →→ å påvirke ↓	*Blære eller urin*	*Tarm eller endetarm*	*Skjede eller bekkenbunnen*
1. din evne til å gjøre husarbeid (matlaging, rengjøring, klesvask)?	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye
2. din fysiske aktivitet, som turgåing, svømming eller annen mosjon?	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye
3. dine fritidsaktiviteter som å gå på kino eller konsert?	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye
4. din mulighet til å reise med bil eller buss i mer enn 30 minutter hjemmefra?	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye
5. din deltakelse i sosiale aktiviteter utenfor hjemmet?	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye
6. din psykiske helsetilstand (nervøsitet, depresjon osv.)?	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye
7. din følelse av frustrasjon?	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye	☐ Ikke i det hele tatt ☐ Litt ☐ I noen grad ☐ Ganske mye

References

1. Abrams P, Andersson KE, Birder L, et al. Fourth International Consultation on Incontinence Recommendations of the International Scientific Committee: evaluation and treatment of urinary incontinence, pelvic organ prolapse, and fecal incontinence. Neurourol Urodyn. 2010;29:213–240. doi:10.1002/nau.20870.
2. Nygaard I, Barber MD, Burgio KL, et al. Prevalence of symptomatic pelvic floor disorders in US women. JAMA. 2008;300:1311–1316. doi:10.1001/jama.300.11.1311.
3. Barber MD, Kuchibhatla MN, Pieper CF, Bump RC. Psychometric evaluation of 2 comprehensive condition-specific quality of life instruments for women with pelvic floor disorders. Am J Obstet Gynecol. 2001;185:1388–1395. doi:10.1067/mob.2001.118659.
4. Barber MD, Chen Z, Lukacz E, et al. Further validation of the short form versions of the Pelvic Floor Distress Inventory (PFDI) and Pelvic Floor Impact Questionnaire (PFIQ). Neurourol Urodyn. 2011;30:541–546. doi:10.1002/nau.20934.
5. Barber MD, Walters MD, Bump RC. Short forms of two condition-specific quality-of-life questionnaires for women with pelvic floor disorders (PFDI-20 and PFIQ-7). Am J Obstet Gynecol. 2005;193:103–113. doi:10.1016/j.ajog.2004.12.025.
6. Abrams P, Cardozo L, Khoury S, Wein A. Incontinence. 5th ed. Paris: Health Publication; 2013.

7. Teleman P, Stenzelius K, Iorizzo L, Jakobsson U. Validation of the Swedish short forms of the Pelvic Floor Impact Questionnaire (PFIQ-7), Pelvic Floor Distress Inventory (PFDI-20) and Pelvic Organ Prolapse/and Urinary Incontinence Sexual Questionnaire (PISQ-12). Acta Obstet Gynecol Scand. 2011;90:483–487. doi:10.1111/j.1600-0412.2011.01085.x.
8. Kaplan PB, Sut N, Sut HK. Validation, cultural adaptation and responsiveness of two pelvic-floor-specific quality-of-life questionnaires, PFDI-20 and PFIQ-7, in a Turkish population. Eur J Obstet Gynecol Reprod Biol. 2012;162:229–233. doi:10.1016/j.ejogrb.2012.03.004.
9. Utomo E, Blok BF, Steensma AB, Korfage IJ. Validation of the Pelvic Floor Distress Inventory (PFDI-20) and Pelvic Floor Impact Questionnaire (PFIQ-7) in a Dutch population. Int Urogynecol J. 2014;25:531–544. doi:10.1007/s00192-013-2263-z.
10. Dewolf L, Koller M, Velikova G, et al. EORTC Quality of Life Group Translation Procedure. 3rd ed. Brussels: EORTC; 2009.
11. Dalkey NC. The Delphi method: an experimental study of group opinion. Santa Monica: Rand Corporation; 1969.
12. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. PLoS One. 2011;6, e20476. doi:10.1371/journal.pone.0020476.
13. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine. 2000;25:3186–3191. doi:10.1097/00007632-200012150-00014.
14. Bump R, Mattiasson A, Bø K, Brubaker L, DeLancey JOL, Klarskov P, et al. The standardization of terminology of female pelvic organ prolapse and pelvic floor dysfunction. Am J Obstet Gynecol. 1996;175:10–17. doi:10.1016/S0002-9378(96)70243-0.
15. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60:34–42. doi:10.1016/j.jclinepi.2006.03.012.
16. Gandek B, Ware Jr JE, Aaronson NK, et al. Tests of data quality, scaling assumptions, and reliability of the SF-36 in eleven countries: results from the IQOLA Project. J Clin Epidemiol. 1998;51:1149–1158. doi:10.1016/S0895-4356(98)00106-1.
17. Maruish ME, editor. User's manual for the SF-36v2 Health Survey. 3rd ed. Lincoln: QualityMetric; 2001.
18. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: a practical guide. Cambridge: Cambridge University Press; 2011.
19. 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;19:539–549. doi:10.1007/s11136-010-9606-8.
20. 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;63:737–745. doi:10.1016/j.jclinepi.2010.02.006.
21. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 2nd ed. Oxford: Oxford University Press; 1995.
22. Gelhorn HL, Coyne KS, Sikirica V, Gauld J, Murphy M. Psychometric evaluation of health-related quality-of-life measures after pelvic organ prolapse surgery. Female Pelvic Med Reconstr Surg. 2012;18:221–226. doi:10.1097/SPV.0b013e31825e6707.
23. Srikrishna S, Robinson D, Cardozo L. Validation of the Patient Global Impression of Improvement (PGI-I) for urogenital prolapse. Int Urogynecol J. 2010;21:523–528. doi:10.1007/s00192-009-1069-5.
24. Due U, Brostrøm S, Lose G. Validation of the Pelvic Floor Distress Inventory-20 and the Pelvic Floor Impact Questionnaire-7 in Danish women with pelvic organ prolapse. Acta Obstet Gynecol Scand. 2013;92:1041–1048. doi:10.1111/aogs.12189.
25. Rogers RG, Rockwood TH, Constantine ML, et al. A new measure of sexual function in women with pelvic floor disorders (PFD): the Pelvic Organ Prolapse/Incontinence Sexual Questionnaire, IUGA-Revised (PISQ-IR). Int Urogynecol J. 2013;24:1091–1103. doi:10.1007/s00192-012-2020-8.
26. Handa VL, Barber MD, Young SB, et al. Paper versus web-based administration of the Pelvic Floor Distress Inventory 20 and Pelvic Floor Impact Questionnaire 7. Int Urogynecol J. 2008;19:1331–1335. doi:10.1007/s00192-008-0651-6.