The Pelvic Girdle Questionnaire: Responsiveness and Minimal Important Change in Women With Pregnancy-Related Pelvic Girdle Pain, Low Back Pain, or Both

Britt Stuge, Hanne Krogstad Jenssen, Margreth Grotle

Background. The Pelvic Girdle Questionnaire (PGQ) is a condition-specific measure for women with pelvic girdle pain (PGP). The PGQ includes items relating to activity/participation and bodily symptoms and has reliability, validity, and feasibility for use in research and clinical practice.

Objective. The purposes of this study were to examine the responsiveness of the PGQ, to determine the minimal important change (MIC) for the PGQ, and to compare the PGQ with other outcome measures.

Design. This study used a prospective cohort design.

Methods. A total of 801 women responded to a booklet of questionnaires in the last trimester of their pregnancy and within 3 months post partum. Responsiveness analyses followed recommendations from the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) checklist. The responsiveness of the PGQ was tested by examining correlations between the change scores of the total PGQ and the other patient-reported outcome measures.

Results. A total of 606 women (76%) reported PGP, low back pain, or both. Of these women, 441 (73%) responded to the follow-up questionnaire post partum. The PGQ (both subscale and total scores) discriminated most accurately between participants who improved and those who did not improve, with an area under the receiver operator characteristic curve of 72%. The MIC values indicated that a change score smaller than 25 for the total score and activity subscale score and a change score of 20 for the symptom subscale score should be regarded as insignificant. Baseline PGQ scores had a large impact on the MIC estimates for the absolute change scores but not on the relative percentage change scores. Five of 6 hypotheses were supported (83%).

Limitations. The type of anchor and definition of important change used might be weaknesses in women whose status is changing from pregnant to post partum.

Conclusions. The PGQ showed acceptable responsiveness in women with PGP, low back pain, or both.

B. Stuge, PT, PhD, Division of Orthopaedic Surgery, Oslo University Hospital, Kirkeveien 166, NO-0407 Oslo, Norway. Address all correspondence to Dr Stuge at: britt.stuge@medisin.uio.no.

H. Krogstad Jenssen, PT, MSc, Department of Physical Medicine and Rehabilitation, Oslo University Hospital.

M. Grotle, PT, PhD, Institute of Physiotherapy, Faculty of Health Sciences, Oslo and Akershus University College of Applied Sciences, Oslo, Norway, and Communication Unit for Musculoskeletal Disorders, Oslo University Hospital.

[Stuge B, Jenssen HK, Grotle M. The Pelvic Girdle Questionnaire: responsiveness and minimal important change in women with pregnancy-related pelvic girdle pain, low back pain, or both. *Phys Ther.* 2017;97:1103–1113.]

Published Ahead of Print: July 29, 2017
Accepted: July 27, 2017
Submitted: December 13, 2016

Pregnancy-related pelvic girdle pain (PGP) is a significant problem worldwide.[1-4] Pelvic girdle pain is described as a disorder which differs from low back pain (LBP) and other musculoskeletal ailments, with a unique clinical presentation requiring specific management.[5-9] Pelvic girdle pain often arises in relation to pregnancy and the pain is related to the pelvic musculoskeletal system. It does not derive from gynecological or urological disorders. To be able to differentiate between PGP and LBP a clinical examination is needed,[5] hence many studies do not separate PGP from LBP.[10,11] Even though no gold standard for diagnosing PGP exists, typically the posterior pelvic pain provocation test is positive in patients with PGP and not in patients with LBP.[5] Although similar and overlapping features may be ascribed to both LBP and PGP, PGP appears to have more impact on disability during pregnancy.[12,13] Although women with LBP show minor problems with daily activities women with PGP commonly may be severely afflicted. The majority of women with PGP recover spontaneously soon after delivery, but about 20% report pain which persists for years.[14,15] Pelvic girdle pain limits most daily activities, the woman's ability to work and is associated with decreased health-related quality of life.[16-19]

To obtain information regarding the impact of PGP on general functioning, or treatment effects, clinicians and researchers must rely on patients' reporting of symptoms and disability, consequently a suitable, reliable and valid outcome measure is required. The Pelvic Girdle Questionnaire (PGQ) is a recent condition-specific outcome measure developed for patients with PGP.[20] The questionnaire includes items relating to activity/participation and bodily symptoms and is suitable for both pregnant and postpartum women with PGP. The PGQ has acceptably high reliability and validity, satisfactory discriminative validity, and it is simple to administer and is feasible for use in both research and clinical practice.[21] The questionnaire is recently recommended as primary outcome for future clinical trials and for application in different cultural settings following proper translation and assessment of cross-cultural equivalence.[22-25]

Patient-reported outcome measures (PROMs) are increasingly used in physical therapy, both by clinicians and researchers.[26] Patient-reported outcome measures capture patients' own opinions on the impact of their condition and enable physical therapists to demonstrate measurable improvements in the clinical outcome of their interventions. A common goal when providing physical therapy to patients with PGP is pain reduction and improved function. To accurately measure change in symptoms and activity limitations there is a need for measurement tools that show responsiveness and are able to detect a minimal important change (MIC) in performance over time.[27] Interpretability, commonly assessed by the MIC of the instrument, is an important aspect of an outcome measurement instrument.[28] The responsiveness and MIC of the PGQ have not been reported.

Hence, the primary aim of this study was to examine the responsiveness of the PGQ and to determine the MIC of the questionnaire. The secondary aim was to compare the responsiveness and MIC of the PGQ with those of other outcome measures commonly used among women with pregnancy-related PGP, LBP, or both.

Methods

Study Participants and Procedure

Participants for this prospective cohort study of pregnant women with a follow-up after delivery were recruited at 4 maternity care centers in Oslo, Norway, from 2013 to 2015. Women in the last trimester of their pregnancy, who could speak and understand Norwegian, were invited to participate by a midwife at each maternity care center. The 4 centers were chosen to be socioeconomically representative of the population in Oslo.

Eligible women who were willing to participate gave written informed consent and self-reported questionnaire data were successively collected from 801. Within 3 months after delivery the women responded to a similar questionnaire sent by mail. The questionnaire consisted of sociodemographic variables (eg, age, weight, height, pregnancy status, parity, education, occupation, and financial status), questions about pain concern, and beliefs in persistence of pain postpartum. The presence and severity of pain were addressed with 2 questions: "In the past 4 weeks, have you had pain in your pelvis and/or low back?" and "If yes, was the pain bad enough to limit your usual activities or change your daily routine for more than 1 day?"[29] A small diagram illustrating the pelvis and the low back accompanied the first question. They were also asked about frequency and location of pain. To perform a clinical examination was not feasible in this large cohort study, hence we had to rely on the women's self-report of PGP and/or LBP. Based on clinical examination it has previously been found that even though some women present with a combination of PGP and LBP, PGP was significantly more prevalent than LBP by late pregnancy.[11] As we know that the prevalence and impact of PGP increases from early to late pregnancy and declines significantly 12 weeks postpartum, we included women during last trimester and within 3 months after delivery.[30,31]

The study was approved by the Regional Committee for Medical Research Ethics in Norway 2012/1626(/REK sør-øst B).

Outcome Measures

To examine different domains the following commonly used PROMs for PGP and LBP were included in the booklet of questionnaires: PGQ,[20] Oswestry Disability Index (ODI) (version 2.0)[32] Disability Rating Index (DRI),[33] 8-Item Short-Form Health Survey (SF-8) for health-related quality of life,[34] the EuroQol Five Dimensions Questionnaire and visual analog scale (EQ-5D-VAS),[35] and questions on PGP and/or LBP intensity in the evening (numeric rating scale; 0–10). The PGQ is a condition-specific instrument that assesses activity limitations (PGQ activity with 20 items) and symptoms (PGQ symptom with 5 items) in women with PGP during pregnancy as well as postpartum.[20] Items are scored on a 4-point descriptive scale (0–3). To

calculate the total PGQ score, all scores are summarized and divided by the total possible score of 75 (maximum possible scores of 60 for activity and 15 for symptoms), subsequently recalculated to a percentage, resulting in percentage scores ranging from 0 (no disability) to 100 (severe disability). If the respondents choose the option "not applicable" to an item, 3 points should be deducted from the total possible score. The ODI and DRI were included as they are frequently used to assess disability and function, the numeric rating scale was used to examine pain intensity, and the SF-8 and EQ-5D-VAS were used to assess quality of life.[9,13,18,20]

Responsiveness and External Anchor

Responsiveness is defined as the ability of an instrument to detect change over time in the construct to be measured.[36] When analyzing the responsiveness and interpretability of the instruments we followed the definitions and recommendations from the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) checklist.[37] First, predefined hypotheses were tested and the details are described under the Statistical analysis section. Second, responsiveness was tested by investigating discriminative ability between participants who improved and those who did not improve, based on an external anchor to identify whether participants have changed over time or not. The Self-Rated Global Perceived Effect (GPE) was used as the anchor (gold standard) with the following question: How is your PGP and/or LBP now compared with late pregnancy? The response categories included (1) "completely recovered," (2) "much improved," (3) "slightly improved," (4) "unchanged," (5) "little worse," (6) "much worse," and (7) "extremely worse."[38] Participants who responded "completely recovered" and "much improved" were classified as having achieved a clinically important improvement, whereas responses 3 through 7 were classified as not having improved. The category "slightly improved" was not considered important enough because it is to be expected that PGP and LBP should disappear postpartum[39] and to avoid socially desirable answers.

Data Analysis

Statistical analysis was performed by use of IBM SPSS 23.0 (SPSS Inc, Chicago, Illinois). Descriptive statistics were used to analyze participant characteristics and to score distributions in the outcome measures. Continuous variables were presented as means with standard deviation in case of normal distribution, or medians with inter quartile range for the rest. Categorical variables were presented as frequencies with percentages. Potential floor and ceiling effects were explored by analyzing the number of people scoring the highest and lowest score on the outcome measures. A criterion of 15% to define floor and ceiling effect was used.[40] Only participants who had experienced PGP and/or LBP during pregnancy and those with complete baseline and postpartum scores (responders) for each outcome measure were included in the responsiveness analyses. The absolute change scores in the PGQ, ODI, DRI, and evening pain were calculated as baseline scores minus postpartum scores, whereas the absolute change scores in the SF-8 subscales were calculated the opposite direction (due to scale direction), so an increasing positive change score equals improvement in the health condition.

In order to evaluate responsiveness of the PGQ, the following hypotheses were tested by Spearman's correlation coefficients (rho) of change values (postpartum – baseline scores) between the total PGQ and ODI, DRI, EQ-5D-VAS, evening pain, SF-8 physical subscale, and SF-8 mental subscale. All the change scores were calculated so the larger positive change scores, the larger improvement. We expected a high positive correlation coefficient between the total PGQ change scores and the change scores of the ODI, DRI, and SF-8 physical, as they all measure physical function. Furthermore, we expected a moderate positive correlation coefficient between the total PGQ change scores and the change scores of the EQ-5D-VAS and evening pain, as they measure different constructs (health-related quality of life and pain). For the SF-8 mental scale, which also assesses a different construct (mental functioning), we expected to have a low correlation with the total PGQ change scores. Coefficients of greater than or equal to .60, less than .60 to greater than .30, and less than or equal to .30 were defined as high, moderate, and low correlations, respectively.[41] Adequate construct validity was established if 75% or more of the correlations corresponded to the predefined hypothesis.[41]

We analyzed the area under the receiver operating characteristic (ROC) curves for the change scores of all the outcome instruments by using dichotomization of the participants' global scores. Sensitivity was defined as the proportion of participants who were correctly classified as having improved (categories 1 and 2), and specificity was defined as the proportion of participants who were correctly classified as not having improved (categories 3 to 7). In the ROC curve, every possible change score from baseline to postpartum was plotted for all instruments using the dichotomized global score as an anchor. The area under the curve (AUC) corresponds with the possibility of correctly predicting a participant's improvement when this is really the case, and reflects how responsive the instruments are in detecting a change in the underlying construct. An AUC of at least 0.70 is considered to be adequate.[37]

The MIC was the smallest change in score in the construct to be measured that participants considered important[36] and was calculated on the basis of the sensitivity and specificity results from the ROC analysis described above. The cutoff point of the change scores with the fewest misclassified participants was used. This corresponds to the upper left point on the ROC curve and can be interpreted as the point or value that yields the lowest overall misclassification.[42] To explore whether the AUC and MIC estimates was dependent on baseline severity, additional analysis was performed. The AUC and the estimate for MIC (with sensitivity and specificity values) were calculated for participants who scored in the lower 25th percentile, the middle 50th percentile, and the upper 75th percentile on the baseline PGQ (both subscale and total scores). Furthermore,

the relative percentage change scores for the PGQ, which takes into account the baseline level, were checked, both for the total material and for the different levels of baseline scores. Finally, for interpretability of the MIC values, they were compared to the smallest detectable change from a previous validation study of the PGQ.[21]

Results

The 801 women had a mean age of 31 years (SD = 4) and were, on average, in week 34 of pregnancy (SD = 2); 74% were pregnant for the first time. Of the 801 women, 606 women (76%) reported PGP and/or LBP; the remaining 195 reported neither. There were no statistically significant differences between the groups regarding gestational week, body mass index, parity, education, and occupation. The women with PGP and/or LBP were marginally younger and 51% of the women with pain were on sick leave compared with 35% of the remainder (P<.001). Women without pain managed slightly better financially (P = .042). Of the 606 women with PGP and/or LBP, 441 (73%) responded to the follow-up questionnaire postpartum (Tab. 1).

Scores on the Outcome Measures: Missing and Ceiling and Floor Effects

Among the responding women with PGP and/or LBP, there were few missing data in the outcome measures administered, both at baseline by late pregnancy and at follow-up postpartum. The PGQ showed the least missing data. In general, few of the participants scored the lowest and highest values, except for the pain concern item, in which 19.6% reported no pain concern at baseline. None of the outcomes showed a ceiling effect (highest possible score). The women seemed to be moderately affected by PGP and/or LBP at baseline (Tab. 1).

Self-Rated GPE Scores and Baseline to Postpartum Change Scores

Postpartum, 163 of the responders (37.2%) reported complete recovery, 177 (40.4%) were much improved, 39 (8.9%) were somewhat improved, 7 (1.6%) were the same, and 25 (5.7%) reported worse symptoms. Twenty-seven participants (6.8%) did not respond to the GPE scale postpartum, and 3 had missing values. These participants were excluded from further evaluation; therefore, 411 participants were included in the responsiveness analyses.

Mean late pregnancy and postpartum scores for the outcome measures for each GPE category are presented in Table 2. At baseline, participants who reported complete recovery or worse symptoms tended to report lower severity on the outcome measures than participants in the other categories of the GPE scale. The postpartum follow-up scores corresponded well to a low score (little severity) for those who reported complete recovery or much improved, and slightly higher scores for those who reported a slight improvement. There were unexpectedly high postpartum scores for those who reported no change, but these constituted only 7 cases (1.6% of the sample) with a large SD (Tab. 2).

Expectations regarding the hypotheses of the correlations between change scores are shown in Table 3. The predefined hypotheses were ascertained for 5 out of 6 hypotheses (83%). The change scores of the total PGQ and evening pain correlated higher than expected (rho = .67) (Tab. 3).

ROC Curves and MICs

The PGQ (both subscale and total scores) and evening pain showed most accuracy in correctly discriminating between participants who improved and those who did not improve, with an area under the ROC curve above 70% (Tab. 4). The SF-8 mental and the EQ-5D-VAS showed the lowest discriminative ability with an area under the ROC curve below 60%. The ROC curves are presented in the Figure. The minimal important change values defined as the optimal cutoff point of change scores plotted on the ROC curve are presented in Table 4. The PGQ symptom subscale and the evening pain score had the largest accuracy values. The minimal important change values for the PGQ subscale and total scores were larger than the minimal detectable change values at group level.

Subgroup Analyses of the PGQ

The AUC and MIC estimates for the subgroups with low (<28), middle (28-63), and high (≥63) levels of baseline PGQ scores are presented in Table 5. The AUC estimates improved slightly for all the subgroups when compared to the total sample, whereas the MIC estimates showed a large variation depending on the baseline score. When comparing the general MIC of 25 points for the total PGQ score, the MIC estimate for those scoring in the lower end of the scale (<25th percentile of the total PGQ) was only 6, whereas the MIC for those scored in the middle and higher (>75th) percentile groups were 31 and 39, respectively (Tab. 5).

The percentage change score, which takes the baseline score into account, was also checked for the PGQ. The precision estimates improved when using percentage change score: the AUC increased to 0.87 for the total PGQ score, and the MIC estimate for the percentage change score varied between 50% change (sensitivity = 0.88, specificity = 0.68) and 60% change (sensitivity = 0.83, specificity = 0.80). When investigating the MIC estimates for the subgroups with different levels of baseline PGQ scores, the optimal cutoff for the lower score subgroup was 46% (sensitivity = 0.81, specificity = 0.82), 60% for the medium interval (sensitivity = 0.87, specificity = 0.69), and 55% for the high score subgroup (sensitivity = 0.80, specificity = 0.74).

Discussion

This study examined the responsiveness and interpretability of the PGQ and other commonly used outcome measures in women with PGP and/or LBP during pregnancy and after delivery. The PGQ showed to be a responsive instrument both according to the correlation approach and the ROC approach. Most of the hypotheses concerning correlation between changes in the total PGQ and the other PROMs in this study were supported. Furthermore, the total PGQ scale and the activity and symptom

Table 1.
Baseline Characteristics of Women with Pelvic Girdle Pain (PGP) and/or Low Back Pain (LBP) and Comparison of Questionnaire Responders and Nonresponders (n = 606)[a]

Characteristic	Responders n = 441	Nonresponders n = 165	*P* Value
Age, mean (SD)	31.3 (3.9)	31.0 (4.0)	.34
Gestational weeks, mean (SD)	34.3 (2.2)	34.6 (2.3)	.06
Body mass index, mean (SD)	27.7 (3.3)	27.0 (4.0)	.59
Primipara, no. (%)	327 (74)	116 (70)	.36
Education, no. (%)			.51
Low education level (<12 y)	25 (15)	57 (13)	
High education level (≥12 y)	384 (87)	140 (85)	
Born in Norway, no. (%)	370 (84)	131 (79)	.19
Occupation, no. (%)			.29
Fully paid employment	355 (81)	125 (76)	
Partly paid employment	17 (4)	13 (8)	
Parental leave	58 (13)	23 (14)	
Not in paid employment	10 (2)	3 (2)	
On sick leave, no. (%)	211 (51)	99 (63)	.01
Manage financially well, no. (%)			.04
Very well	211 (48)	64 (39)	
Quite well	205 (47)	86 (52)	
Neither well nor badly	22 (5)	15 (9)	
Limited daily activities, no. (%)	239 (54)	94 (57)	.58
Frequency of PGP/LBP, no. (%)			.57
Some days	203 (46)	69 (42)	
Most days	97 (22)	42 (26)	
Every day	140 (32)	54 (33)	
Location of PGP/LBP, no. (%)			
Low back	258 (65)	87 (63)	.68
Pelvis	341 (81)	137 (87)	.11
Location of PGP, no. (%)			
Front (symphysis)	218 (55)	89 (57)	.71
Right sacroiliac joint	155 (39)	58 (37)	.70
Left sacroiliac joint	150 (38)	62 (40)	.70
Over sacrum	153 (38)	57 (36)	.70
Pain concern,b mean (SD)	4.5 (2.2)	4.9 (2.1)	.04
Belief in persistence of pain postpartum, no. (%)	32 (7)	10 (6)	.64
Evening PGP/LBP intensity,[b] mean (SD)	4.5 (2.2)	4.9 (2.1)	.04
PGQ activity subscale score, mean (SD)	42.9 (22.4)	45.8 (22.1)	.16
PGQ symptom subscale score, mean (SD)	43.4 (23.2)	48.7 (23.7)	.01
Total PGQ score, mean (SD)	43.0 (22.0)	46.4 (21.8)	.09
ODI score, mean (SD)	22.4 (14.4)	15.3 (25.2)	.04
DRI score, mean (SD)	38.3 (21.5)	42.8 (20.8)	.02
SF-8 physical subscale score, mean (SD)	39.7 (9.9)	38.0 (9.9)	.05
SF-8 mental subscale score, mean (SD)	48.6 (8.0)	45.4 (9.6)	.001
EQ-5D-VAS score, mean (SD)	74.6 (17.0)	71.1 (18.6)	.031

[a] DRI = Disability Rating Index, EQ-5D-VAS = EuroQol Five Dimensions Questionnaire and visual analog scale, NRS = numeric rating scale, ODI = Oswestry Disability Index, PGQ = Pelvic Girdle Questionnaire, SF-8 = 8-Item Short-Form Health Survey.
[b] Determined with the NRS (from 0 to 10).

Table 2.
Mean Change Scores (SD) for Outcomes According to Categories in the Global Perceived Effect (GPE) Scale (n = 411)[a]

Outcome	Baseline Late Pregnancy Mean (SD)	Postpartum Mean (SD)	Change Mean (SD)
PGQ activity subscale (0–100)			
Completely recovered	34.8 (20.8)	2.0 (4.5)	32.8 (21.5)
Much improved	51.0 (19.3)	15.3 (13.6)	35.8 (18.6)
Slightly improved	52.5 (23.5)	30.6 (17.7)	21.9 (18.9)
No change	53.7 (26.7)	27.8 (28.9)	25.9 (16.7)
Worse (a little, much, more than ever)	38.6 (23.0)	29.5 (12.3)	9.1 (21.8)
PGQ symptom subscale (0–100)			
Completely recovered	35.4 (21.3)	1.2 (5.0)	34.1 (22.0)
Much improved	52.3 (20.8)	14.6 (15.2)	37.7 (21.5)
Slightly improved	50.6 (23.6)	28.4 (19.4)	22.3 (25.5)
No change	49.5 (26.9)	37.1 (28.0)	12.4 (14.1)
Worse (a little, much, more than ever)	41.1 (23.1)	32.8 (17.8)	8.3 (22.6)
Total PGQ (0–100)			
Completely recovered	35.0 (20.4)	1.8 (4.3)	33.0 (21.0)
Much improved	51.3 (18.9)	15.1 (13.3)	36.2 (18.4)
Slightly improved	52.2 (22.9)	30.2 (17.3)	22.0 (19.2)
No change	52.8 (26.6)	29.7 (28.5)	23.1 (15.6)
Worse (a little, much, more than ever)	39.1 (22.6)	30.1 (12.8)	9.0 (21.1)
ODI (0–100)			
Completely recovered	18.2 (12.5)	1.0 (3.1)	17.1 (12.9)
Much improved	26.9 (13.9)	8.0 (7.1)	18.8 (12.5)
Slightly improved	28.0 (16.0)	16.1 (12.8)	12.0 (12.8)
No change	28.0 (23.0)	15.9 (19.5)	12.1 (10.8)
Worse (a little, much, more than ever)	20.7 (13.1)	15.1 (8.2)	5.6 (10.9)
DRI (0–100)			
Completely recovered	32.5 (20.5)	5.6 (7.8)	26.9 (21.7)
Much improved	45.2 (19.7)	15.6 (13.8)	29.6 (19.3)
Slightly improved	44.5 (21.5)	23.5 (17.1)	21.0 (19.2)
No change	45.3 (24.6)	26.5 (28.7)	18.8 (16.8)
Worse (a little, much, more than ever)	33.9 (20.9)	22.7 (15.0)	11.2 (21.8)
EQ-5D-VAS (0–100)			
Completely recovered	78.6 (16.4)	85.0 (13.8)	6.4 (19.9)
Much improved	70.7 (16.8)	79.1 (13.0)	9.0 (20.0)
Slightly improved	71.0 (19.1)	76.3 (11.9)	5.4 (18.5)
No change	69.3 (17.0)	66.4 (12.8)	–2.9 (12.2)
Worse (a little, much, more than ever)	73.0 (16.5)	71.6 (12.6)	–1.4 (17.7)
SF-8 physical subscale (100–0)			
Completely recovered	42.5 (9.1)	54.8 (4.9)	12.3 (10.2)
Much improved	36.5 (9.6)	50.3 (6.2)	13.6 (9.6)
Slightly improved	36.8 (9.6)	46.3 (5.9)	9.4 (8.4)
No change	36.5 (11.2)	41.4 (12.9)	4.9 (7.3)
Worse (a little, much, more than ever)	41.2 (9.8)	44.3 (5.8)	2.4 (10.0)
SF-8 mental subscale (100–0)			
Completely recovered	49.7 (7.4)	51.9 (5.9)	2.2 (8.7)
Much improved	47.1 (8.3)	49.5 (7.1)	2.3 (9.8)
Slightly improved	48.2 (9.5)	49.4 (10.1)	1.4 (9.2)
No change	50.2 (5.9)	48.6 (14.5)	–1.7 (14.8)
Worse (a little, much, more than ever)	49.5 (9.2)	49.7 (8.5)	–0.2 (7.7)
Evening pain (0–10)			
Completely recovered	3.9 (2.0)	0.2 (0.9)	3.7 (2.1)
Much improved	5.2 (2.0)	2.2 (2.1)	2.9 (2.4)
Slightly improved	5.2 (2.2)	3.9 (2.4)	1.4 (1.7)
No change	5.9 (2.5)	4.7 (2.0)	1.1 (2.0)
Worse (a little, much, more than ever)	3.9 (2.2)	3.8 (2.2)	0.1 (2.0)

[a] DRI = Disability Rating Index, EQ-5D-VAS = EuroQol Five Dimensions Questionnaire and visual analog scale, ODI = Oswestry Disability Index, PGQ = Pelvic Girdle Questionnaire, SF-8 = 8-Item Short-Form Health Survey.

Table 3.
Spearman Correlation Coefficients Between the Change Scores of the Total PGQ and the other PROMs, n=411[a] The sample includes those with complete follow-up

Measure	Hypothesis	Change in Total PGQ (rho)	Supported Hypothesis
Change in ODI	High positive correlation (≥.60)	.820	Yes
Change in DRI	High positive correlation (≥.60)	.790	Yes
EQ-5D-VAS	Moderate positive correlation (<.60 and >.30)	.386	Yes
Evening pain	Moderate positive correlation (<.60 and >.30)	.670	No
SF-8 physical subscale	High positive correlation (≥.60)	.668	Yes
SF-8 mental subscale	Low positive correlation (≤.30)	.206	Yes

[a] For all the change scores, a higher score indicates more improvement. DRI = Disability Rating Index, EQ-5D-VAS = EuroQol Five Dimensions Questionnaire and visual analog scale, ODI = Oswestry Disability Index, PGQ = Pelvic Girdle Questionnaire, PROMs = patient-reported outcome measures, SF-8 = 8-Item Short-Form Health Survey.

Table 4.
Tests of Responsiveness of Outcome Measures, n=411[a] The sample includes those with complete follow-up

Measure	AUC (95% CI)	MIC	MIC Sensitivity	MIC Specificity
PGQ (0–100)				
Activity	0.72 (0.65–0.78)	25	0.67	0.61
Symptom	0.72 (0.65–0.79)	20	0.70	0.63
Total	0.72 (0.66–0.79)	25	0.67	0.61
ODI (0–100)	0.68 (0.61–0.74)	14	0.56	0.68
DRI (0–100)	0.65 (0.58–0.72)	23	0.61	0.61
EQ-5D-VAS (0–100)	0.60 (0.53–0.68)	2	0.59	0.56
Evening pain (0–10)	0.80 (0.75–0.86)	1.5	0.80	0.67
SF-8 physical subscale (100–0)	0.67 (0.60–0.74)	10	0.62	0.60
SF-8 mental subscale (100–0)	0.55 (0.48–0.63)	1	0.50	0.48

[a] AUC = area under the receiver operating characteristic curve, DRI = Disability Rating Index, EQ-5D-VAS = EuroQol Five Dimensions Questionnaire and visual analog scale, MIC = minimal important change, ODI = Oswestry Disability Index, PGQ = Pelvic Girdle Questionnaire, SF-8 = 8-Item Short-Form Health Survey.

subscales discriminated between participants who improved and those who did not improve, with an area under the ROC curve of 72%. The MIC values for the total sample indicated that a change score smaller than 25 for total score and activity subscale, and 20 for the symptom subscale, should be regarded as insignificant to the women. However, this study also showed that the MIC estimates varied largely according to the baseline scores, which is important to consider when interpreting MIC estimates; for participants with low baseline PGQ scores a MIC of 6 can discriminate between participants who improved and those who did not improve, whereas for participants with baseline PGQ scores in the middle or higher percentiles, a change score of at least 31 and 39, respectively, was necessary to distinguish between participants who improved and those who did not improve. At the group level, we recommend using percentage change scores, which adjust for baseline scores, as that reduced the differences among the MIC estimates in subgroups with low, middle, and high baseline scores.

A significant strength of this study is the large sample size that allowed for subgroup analyses of severity of PGP and/or LBP, which clearly showed that different MIC values should be used for women with low, middle or high baseline scores. The MIC refers to meaningful change based on the longitudinal within-person scales[43] and MIC values have previously been shown to be highly dependent on baseline values.[44] Another strength is our investigation of the MIC cutoff values for the percentage change scores, which are more stable than absolute change scores.[45] The current study showed that the discriminative ability improved when using percentage change scores for the PGQ and there was a smaller difference in MIC cutoff values between those who scored in the lower and higher percentiles at baseline. Hence, a percentage change score will provide a more accurate classification when interpreting changes for a group of people, whereas an absolute change score might be easier to use on an individual basis.

It is a limitation that our study was conducted in a sample of pregnant women with self-reported PGP and/or LBP. As the women were not clinically examined we cannot decide whether they had PGP or LBP, or a combination. However, more than 80% of the women reported pain located in the pelvis.

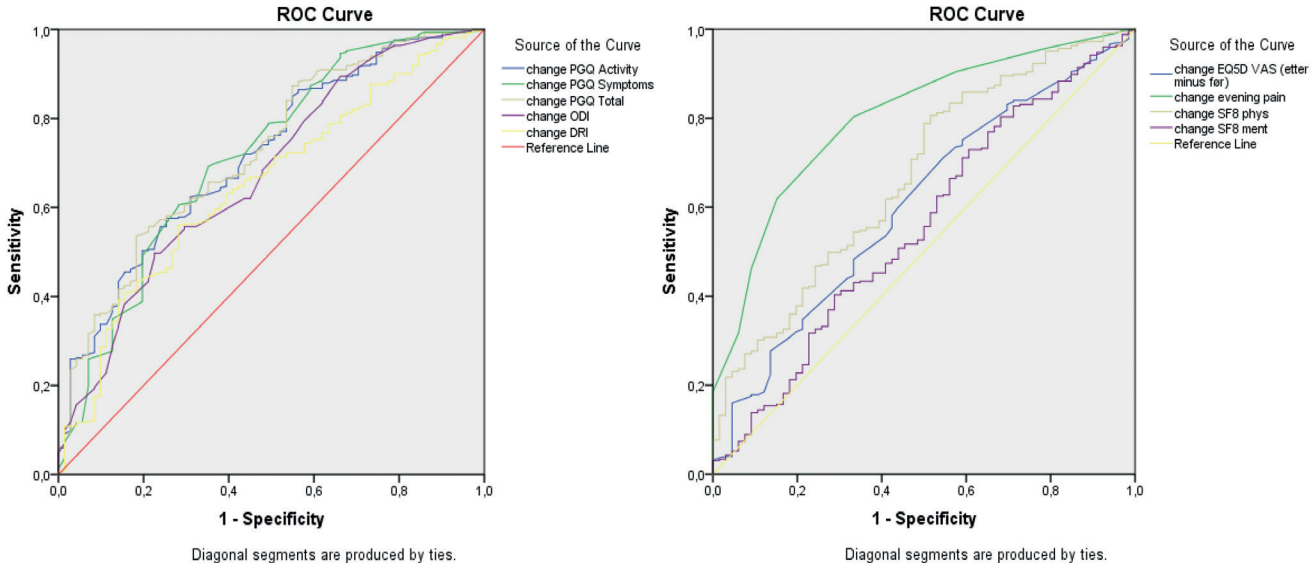

Figure.
Receiver operating characteristic (ROC) curves of key outcomes against self-rated change. DRI = Disability Rating Index, EQ5D VAS = EuroQol Five Dimensions Questionnaire and visual analog scale, ODI = Oswestry Disability Index, PGQ = Pelvic Girdle Questionnaire, SF8 ment = 8-Item Short-Form Health Survey mental subscale, SF8 phys = 8-Item Short-Form Health Survey physical subscale.

In previous studies where pregnant women were clinically examined, PGP was significantly more prevalent than LBP,[3,13] and a recent study showed that self-reported PGP was verified by specific clinical tests in nearly all cases.[25] Most probably women in our study who reported PGP had PGP, with or without concomitant LBP. However, as we did not clinically examine the women we do not know how many did not have PGP. As women with LBP has shown to be significant less afflicted than women with PGP,[13] the influence of including women with LBP might be that they would show lower scores on the PGQ.

A main limitation of the current study, as in most studies of responsiveness, is the lack of an optimal external anchor for change in health status. We used the GPE, despite evidence that this external anchor has many weaknesses, eg, it reflects more the health condition at follow-up than the change in the health condition.[38,46] In the present study we also found that the GPE did not work optimal, in particular among the women who did not improve. The women scoring slightly improved and unchanged showed rather high positive change scores in the instruments assessing physical function. Therefore, it seems like the global change scores might reflect constructs other than changes in pain-related physical function, for example aspects in a changed life situation with a newborn child. The lack of nonoptimal external criteria for this setting demands a careful interpretation of the MIC estimates in this study. The correlational approach for the responsiveness analysis might be more reliable, as this approach does not use an external anchor.[40] Even though an anchor-based method is useful due to the MIC estimates there is no consensus as to which method is best.[47,48] It is a strength that we used both approaches. We also used a domain-specific question related to the anchor and it has recently been shown that a domain-specific anchor is more valid than a generic anchor for MIC calculations.[49]

The anchor-based method requires the choice of a sensible cutoff point of important change. There is debate about whether the category "slightly improved" should be considered important or not. We concluded that it should not, consistent with others' findings that considering slight improvements as important decreases the accuracy in distinguishing participants who improved from those who did not improve.[50] The cutoff for an important change will be dependent on the intervention or the expected course of the disease. The status of our participants changed from late pregnancy to postpartum. Pregnancy-related PGP will often resolve or improve spontaneously after delivery.[39] The large improvement among our participants reporting an improved condition, supports this. Therefore, this material provides important data on normally expected improvement in pregnancy-related PGP in a general pregnant population. This could have clinical implications because the expectation and most likely change in the participants' condition is a reduction in their functional disability level over time, and associated reduction of the total PGQ score after delivery. Simultaneously, the activity required as the mother to a newborn is different from being pregnant. Recall-bias, postpartum memory and expectations might have influenced the scoring on the GPE. Hence, the ratings might not really have taken into account the

Table 5.
Mean (SD) of Pelvic Girdle Questionnaire (PGQ) Baseline Scores and Change Scores

Parameter	Value for the Following Subgroup:[a]		
	Low (n = 106)	Middle (n = 207)	High (n = 98)
PGQ activity subscale			
Baseline score mean (SD)	15.5 (8.1)	45.1 (10.2)	72.8 (7.1)
Change score mean (SD)	8.3 (12.2)	35.2 (14.3)	48.6 (19.2)
AUC[b]	0.83 (0.74 to 0.92)	0.76 (0.66 to 0.87)	0.86 (0.78 to 0.94)
MIC[c]	6 (0.71 and 0.74)	32 (0.73 and 0.72)	34 (0.87 and 0.70)
PGQ symptom subscale			
Baseline score mean (SD)	18.4 (11.2)	45.4 (14.7)	71.6 (11.3)
Change score mean (SD)	11.1 (15.3)	35.7 (18.4)	49.4 (23.1)
AUC[b]	0.77 (0.63 to 0.90)	0.76 (0.65 to 0.87)	0.81 (0.71 to 0.91)
MIC[c]	7 (0.82 and 0.63)	27 (0.74 and 0.72)	40 (0.74 and 0.70)
Total PGQ			
Baseline score mean (SD)	16.1 (7.6)	45.1 (9.9)	72.5 (6.5)
Change score mean (SD)	8.9 (11.8)	35.3 (14.2)	48.7 (19.1)
AUC[b]	0.85 (0.76 to 0.93)	0.79 (0.69 to 0.89)	0.86 (0.78 to 0.94)
MIC[c]	6 (0.72 and 0.74)	31 (0.74 and 0.72)	39 (0.85 and 0.70)

[a] Low = low (<28) baseline PGQ scores, Middle = middle (28–62) baseline PGQ scores, High = high (>62) baseline PGQ scores.
[b] Area under the receiver operating characteristic curve (AUC) with 95% confidence intervals.
[c] Estimate for minimal important change (MIC) for participants scoring in the lower 25 percentile, the middle 50 percentile, and the upper 75 percentile of the baseline scores of the Pelvic Girdle Questionnaire (PGQ). Values in parentheses are sensitivity and specificity.

change in health status but rather been influenced by that status at time of assessment.[38] It might also be that the amount of improvement in a sample of pregnant women seeking treatment for their PGP during pregnancy would be less, as severity of PGP during pregnancy has been found to be a predictor for recovery postpartum.[39]

Clinical trials in which people undergo therapy and pre- and post-treatment assessment are commonly used as comparator instruments in the assessment of responsiveness.[44,51] This study followed the natural course of a cohort of pregnant women until after delivery, without any treatment being offered. There is however evidence that the MIC of an instrument is not substantially influenced by an intervention or different types of interventions.[45,44] As the women were recruited from maternity care centers, they may well have been less afflicted by PGP and/or LBP than a sample of people seeking treatment. The current cohort might however be considered representative of a general pregnant population, and the findings are therefore useful as reference or norm values for future epidemiological studies.

The PGQ is a newly developed questionnaire,[20] and no previous studies on its responsiveness have been published; therefore, no comparisons can be made. According to our findings, the PGQ is able to detect "real" changes in symptoms and activity for women with PGP. Furthermore, this study showed that the MICs were larger than the measurement error and estimates of smallest detectable change of the PGQ.[21] Since the smallest detectable change estimate for the PGQ was developed in a clinical sample, and our current material represents the general population, we need to be careful with a direct comparison. A test-retest study from a similar population as in the present study should be carried out in order to be sure of this comparison. We conclude that it is likely that the PGQ is able to distinguish measurement error from the MIC, making the interpretability of the change scores possible. The PGQ also showed excellent responsiveness when using the correlational approach by comparing the PGQ change scores with the change scores of other PROMs.

Responsiveness and MIC of an instrument are often population and context specific,[40] which should be taken into account before generalizing to other populations. Measurement properties of the PGQ have so far been examined in Norwegian and Spanish samples[20,21,25] and need to be examined in other cultures and languages.[24] It is necessary to determine whether the clinometric performance of the items is consistent across different languages and cultures and whether it is an adequate reflection of the performance of the original instrument.[36] Furthermore, to increase the knowledge on GPE as a valid indicator of change in health status, future methodological studies should investigate sets of different anchors to better understand which might provide the most reliable and valid assessment.[38,46]

In conclusion, for women with PGP and/or LBP, the PGQ and evening pain showed

acceptable responsiveness and ability to discriminate between women who improved and those who did not improve.

Author Contributions and Acknowledgments

Concept/idea/research design: B. Stuge, H.K. Jenssen, M. Grotle
Writing: B. Stuge, H.K. Jenssen, M. Grotle
Data collection: B. Stuge, H.K. Jenssen
Data analysis: B. Stuge, H.K. Jenssen, M. Grotle
Project management: B. Stuge, H.K. Jenssen
Fund procurement: B. Stuge
Providing participants: B. Stuge, H.K. Jenssen
Providing facilities/equipment: B. Stuge
Providing institutional liaisons: B. Stuge
Consultation (including review of manuscript before submitting): B. Stuge

The authors thank the staff who gathered data at maternity care centers Sagene, Frogner, Nordre Aker, and Østensjø in Oslo, Norway.

Ethics Approval

The study was approved by the Regional Committee for Medical Research Ethics in Norway (2012/1626/REK sør-øst B).

Funding

Grant support was provided by The Norwegian Fund for postgraduate training in physical therapy.

Disclosures/Presentations

The authors completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and reported no conflicts of interest.

This article was adapted in part from a presentation given at the 9th Interdisciplinary World Congress on Low Back and Pelvic Girdle Pain, Singapore, October 31, 2016, and a poster presentation at the World Confederation for Physical Therapy in Cape Town, South Africa, July 3, 2017.

DOI: 10.1093/ptj/pzx078

References

1 Bjorklund K, Bergstrom S. Is pelvic pain in pregnancy a welfare complaint? *Acta Obstet Gynecol Scand.* 2000;79:24–30.

2 Charpentier K, Leboucher J, Lawani M, Toumi H, Dumas GA, Pinti A. Back pain during pregnancy and living conditions: a comparison between Beninese and Canadian women. *Ann Phys Rehabil Med.* 2012;55:148–159.

3 Mousavi SJ, Parnianpour M, Vleeming A. Pregnancy related pelvic girdle pain and low back pain in an Iranian population. *Spine (Phila Pa 1976).* 2007;32:E100–E104.

4 Chang HY, Jensen MP, Lai YH. How do pregnant women manage lumbopelvic pain? Pain management and their perceived effectiveness. *J Clin Nurs.* 2015;24:1338–1346.

5 Vleeming A, Albert HB, Ostgaard HC, Sturesson B, Stuge B. European guidelines for the diagnosis and treatment of pelvic girdle pain. *Eur Spine J.* 2008;17:794–819.

6 Pennick V, Liddle SD. Interventions for preventing and treating pelvic and back pain in pregnancy. *Cochrane Database Syst Rev.* 2013;8:CD001139.

7 Gutke A, Ostgaard HC, Oberg B. Pelvic girdle pain and lumbar pain in pregnancy: a cohort study of the consequences in terms of health and functioning. *Spine (Phila Pa).* 2006;31:E149–E155.

8 O'Sullivan PB, Beales DJ. Diagnosis and classification of pelvic girdle pain disorders, part 2: illustration of the utility of a classification system via case studies. *Man Ther.* 2007;12:e1–e12.

9 Stuge B, Lærum E, Kirkesola G, Vøllestad N. The efficacy of a treatment program focusing on specific stabilizing exercises for pelvic girdle pain after pregnancy: a randomized controlled trial. *Spine (Phila Pa 1976).* 2004;29:351–359.

10 van Benten E, Pool J, Mens J, Pool-Goudzwaard A. Recommendations for physical therapists on the treatment of lumbopelvic pain during pregnancy: a systematic review. *J Orthop Sports Phys Ther.* 2014;44:464–415.

11 Liddle SD, Pennick V. Interventions for preventing and treating low-back and pelvic pain during pregnancy. *Cochrane Database Syst Rev.* 2015;9:CD001139.

12 Ostgaard HC, Zetherstrom G, Roos-Hansson E, Svanberg B. Reduction of back and posterior pelvic pain in pregnancy. *Spine (Phila Pa 1976).* 1994;19:894–900.

13 Robinson HS, Mengshoel AM, Bjelland EK, Vollestad NK. Pelvic girdle pain, clinical tests and disability in late pregnancy. *Man Ther.* 2010;15:280–285.

14 Wu WH, Meijer OG, Uegaki K, et al. Pregnancy-related pelvic girdle pain (PPP), I: terminology, clinical presentation, and prevalence. *Eur Spine .J* 2004;13:575–589.

15 Bergstrom C, Persson M, Mogren I. Pregnancy-related low back pain and pelvic girdle pain approximately 14 months after pregnancy: pain status, self-rated health and family situation. *BMC Pregnancy Childbirth.* 2014;14:48.

16 Wormslev M, Juul AM, Marques B, Minck H, Bentzen L, Hansen TM. Clinical examination of pelvic insufficiency during pregnancy: an evaluation of the interobserver variation, the relation between clinical signs and pain and the relation between clinical signs and physical disability. *Scand J Rheumatol.* 1994;23:96–102.

17 Bergstrom C, Persson M, Mogren I. Sick leave and healthcare utilisation in women reporting pregnancy related low back pain and/or pelvic girdle pain at 14 months postpartum. *Chiropr Man Therap.* 2016;24:7.

18 Gutke A, Lundberg M, Ostgaard HC, Oberg B. Impact of postpartum lumbopelvic pain on disability, pain intensity, health-related quality of life, activity level, kinesiophobia, and depressive symptoms. *Eur Spine J.* 2010;20:440–448.

19 Olsson C, Nilsson-Wikmar L. Health-related quality of life and physical ability among pregnant women with and without back pain in late pregnancy. *Acta Obstet Gynecol Scand.* 2004;83:351–357.

20 Stuge B, Garratt A, Krogstad JH, Grotle M. The pelvic girdle questionnaire: a condition-specific instrument for assessing activity limitations and symptoms in people with pelvic girdle pain. *Phys Ther.* 2011;91:1096–1108.

21 Grotle M, Garratt AM, Krogstad JH, Stuge B. Reliability and construct validity of self-report questionnaires for patients with pelvic girdle pain. *Phys Ther.* 2012;92:111–123.

22 Bishop A, Ogollah R, Bartlam B, et al. Evaluating acupuncture and standard care for pregnant women with back pain: the EASE Back pilot randomised controlled trial (ISRCTN49955124). *Pilot Feasibility Stud.* 2016;2:72.

23 Epstein J, Santo RM, Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *J Clin Epidemiol.* 2015;68:435–441.

24 Girard MP, Marchand AA, Stuge B, Ruchat SM, Descarreaux M. Cross-cultural adaptation of the pelvic girdle questionnaire for the French-Canadian population. *J Manipulative Physiol Ther.* 2016;39:494–499.

25 Rejano-Campo M, Ferrer-Pena R, Urraca-Gesto MA, et al. Transcultural adaptation and psychometric validation of a Spanish-language version of the "Pelvic Girdle Questionnaire." *Health Qual Life Outcomes.* 2017;15:30.

26 Kyte DG, Calvert M, van der Wees PJ, ten HR, Tolan S, Hill JC. An introduction to patient-reported outcome measures (PROMs) in physiotherapy. *Physiotherapy.* 2015;101:119–125.

27 Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. *Health Qual Life Outcomes.* 2006;4:70.

28 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.

29 Dionne CE, Dunn KM, Croft PR, et al. A consensus approach toward the standardization of back pain definitions for use in prevalence studies. *Spine (Phila Pa 1976).* 2008;33:95–103.

30 Robinson HS, Veierod MB, Mengshoel AM, Vollestad NK. Pelvic girdle pain - associations between risk factors in early pregnancy and disability or pain intensity in late pregnancy: a prospective cohort study. *BMC Musculoskelet Disord.* 2010;11:91.

31 Robinson HS, Mengshoel AM, Veierod MB, Vollestad N. Pelvic girdle pain: Potential risk factors in pregnancy in relation to disability and pain intensity three months postpartum. *Man Ther.* 2010;15:522–528.

32 Fairbank JC, Davies JB, Couper J, O'Brien JP. The Oswestry Low Back Pain Disability Questionnaire. *Physiotherapy.* 1980;66:271–273.

33 Salèn BA, Spangfort EV, Nygren ÅL, Nordemar R. The Disability Rating Index: an instrument for the assessment of disability in clinical settings. *J Clin Epidemiol.* 1994;47:1423–1434.

34 Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36); I: conceptual framework and item selection. *Med Care.* 1992;30:473–483.

35 Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med.* 2001;33:337–343.

36 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.

37 Mokkink LB, Terwee CB, Patrick DL, et al. *COSMIN Checklist Manual.* Amsterdam, the Netherlands: EMGO Institute for Health and Care Research, VU University Medical Center; 2012.

38 Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ. Global perceived effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol.* 2010;63:760–766.

39 Bjelland EK, Stuge B, Engdahl B, Eberhard-Gran M. The effect of emotional distress on persistent pelvic girdle pain after delivery: a longitudinal population study. *BJOG.* 2013;120:32–40.

40 de Vet HC, Ostelo RW, Terwee CB, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. *Qual Life Res.* 2007;16:131–142.

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

42 van der Roer N, Ostelo RW, Bekkering GE, van Tulder MW, de Vet HC. Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. *Spine (Phila Pa 1976).* 2006;31:578–582.

43 Beaton DE, Bombardier C, Katz JN, et al. Looking for important change/differences in studies of responsiveness. *J Rheumatol.* 2001;28:400–405.

44 de Vet HC, Foumani M, Scholten MA, et al. Minimally important change values of a measurement instrument depend more on baseline values than on the type of intervention. *J Clin Epidemiol.* 2015;68:518–524.

45 Farrar JT, Young JP, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain.* 2001;94:149–158.

46 Grovle L, Haugen AJ, Hasvik E, Natvig B, Brox JI, Grotle M. Patients' ratings of global perceived change during 2 years were strongly influenced by the current health status. *J Clin Epidemiol.* 2014;67:508–515.

47 Mokkink LB, Terwee CB, Knol DL et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol.* 2010;10:22.

48 Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.* 2008;61:102–109.

49 Ward MM, Guthrie LC, Alba M. Domain-specific transition questions demonstrated higher validity than global transition questions as anchors for clinically important improvement. *J Clin Epidemiol.* 2015;68:655–661.

50 Demoulin C, Ostelo R, Knottnerus JA, Smeets RJ. What factors influence the measurement properties of the Roland-Morris disability questionnaire? *Eur J Pain.* 2010;14:200–206.

51 Chiarotto A, Vanti C, Cedraschi C, et al. Responsiveness and minimal important change of the pain self-efficacy questionnaire and short forms in patients with chronic low back pain. *J Pain.* 2016;17:707–718.